

**NOUVELLE MÉTHODE DE PRÉDICTION D'ÉCHEC DE  
MODÈLES PRÉDICTIFS ET DE DÉTECTION DE  
CHANGEMENTS DE RÉGIMES DANS LES SÉRIES  
CHRONOLOGIQUES**

par

Jean Paul Latyr Faye

Mémoire présenté au Département d'informatique  
en vue de l'obtention du grade de maître ès sciences (M.Sc.)

FACULTÉ DES SCIENCES  
UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, 1<sup>er</sup> mars 2021

Le 1<sup>er</sup> mars 2021

Le jury a accepté le mémoire de Jean Paul Latyr Faye dans sa version  
finale

**Membres du jury**

Professeur Shengrui Wang  
Directeur de recherche  
Département d'informatique

Professeur Djemel Ziou  
Membre interne  
Département d'informatique

Professeur Michael Blondin  
Président-rapporteur  
Département d'informatique

# Sommaire

Aujourd'hui, dans le monde de la finance et de l'économie, les décisions sont prises en basant sur les prédictions de certains modèles. Une mauvaise décision due à un échec de modèle prédictif a généralement des conséquences néfastes. Sur le plan financier et économique, on assiste surtout à des pertes énormes. L'échec des modèles prédictifs est souvent une conséquence de changements de régimes dans les marchés financiers. Être en mesure d'anticiper le moment d'échec des modèles prédictifs permettrait de prédire les changements de régime, et par conséquent, pourrait éviter certaines pertes financières. Avec ce prélude établi, il deviendrait plus facile d'anticiper la durée de survie d'un modèle prédictif que de prédire le temps d'échec.

Ce mémoire a pour but de proposer une solution d'anticipation sur la durée de survie des modèles prédictifs. Ainsi, nous avons développé un nouveau *modèle de survie* prédisant la probabilité de survie d'un modèle de prédiction et ceci, en tout moment  $t$ . En prédisant les probabilités de survie, notre modèle aide, d'une autre manière, à anticiper les changements de régime dans les marchés financiers. Après sa mise en oeuvre, le nouveau modèle de survie est expérimenté en utilisant six différents modèles déjà entraînés que sont la moyenne mobile intégrée auto-régressive (ARIMA), le modèle de marche aléatoire, le modèle PROPHET, le modèle hétéroscédasticité conditionnelle autorégressive généralisée (GARCH), le modèle TBATS (trigonométrique saisonnier, transformation Box-Cox, résidus ARMA, tendance et saisonnalité) et le modèle de réseau de neurones pour les séries chronologiques (NNTAR). Ce choix est justifié par le fait que la performance de ces modèles dépend généralement des propriétés des séries temporelles servant à leur entraînement.

Les résultats obtenus démontrent que le modèle de survie est bien en mesure de prédire la probabilité de survie de ces six modèles et donc permet d'anticiper les chan-

## SOMMAIRE

gements de régimes dans les séries financières ayant permis à l’entraînement ces six modèles. En effet, les prédictions obtenues avec le modèle de survie sur un ensemble de test, montrent un comportement similaire à celui du vrai processus de survie obtenu. En outre, nous avons ainsi confirmé ce résultat en mesurant certaines métriques comme le score de Brier censuré (CBS), l’erreur moyenne absolue (MAE) adaptée à l’analyse de survie et l’index de concordance (C-index). Toutes les valeurs de ces métriques démontrent une bonne performance du modèle de survie à prédire la durée de survie des six modèles utilisés dans l’expérimentation du modèle de survie.

**Mots-clés:** Modèles prédictifs ; séries chronologiques ; réseaux de neurones ; changements de régimes ; analyse de survie.

# Remerciements

Mes premiers remerciements vont à Monsieur Shengrui WANG, mon directeur de mémoire qui a bien voulu m'accueillir dans son groupe de recherche. Merci encore Professeur Shengrui WANG, j'ai beaucoup bénéficié de vos profondes connaissances du domaine, de vos conseils et les choses exceptionnelles qui ont mené à bien l'avancement de mes travaux de maîtrise. Je n'ai jamais imaginé dans ma vie de rencontrer un Professeur aussi exceptionnel que vous sur tous les plans. Votre disponibilité, votre patience et votre gentillesse m'ont beaucoup aidé en tant qu'étudiant international et ont renforcé mes souvenirs inoubliables à l'Université de Sherbrooke. Merci infiniment.

Je remercie profondément le Professeur, Michael BLONDIN d'avoir accepté de présider le jury. Je remercie Professeur, Djemel ZIOU d'avoir accepté de faire parti des membres de jury comme évaluateur.

Je remercie tous les membres du Département d'informatique de m'avoir aidé à surmonter certaines situations difficiles.

J'exprime en fin ma gratitude à toute la communauté sénégalaise de l'Université de Sherbrooke.

# Abréviations

**ARCH** Autoregressive Conditional Heteroscedastic; (Auto-Regressif Hétérocédastique Conditionné)

**GARCH** Generalized AutoRegressive Conditional Heteroskedasticity (AutoRegressive Conditional Heteroskedasticity)

**CBS** Score de Brier Censuré; (Censoring Brier Score)

**SVM** Support Vector Machines (Machine à Vecteurs Support)

**NN** Neural Network (Réseaux de neurones)

**MAE** Mean Absolute Error (Moyenne de l'Erreur Absolue)

**AR** Auto-Regressive (Auto-Regressif)

**MASE** Mean Absolute Scaled Erreur (Moyenne de l'Erreur Absolue échelonnée)

# Table des matières

<b>Sommaire</b>	<b>ii</b>
<b>Remerciements</b>	<b>iv</b>
<b>Abréviations</b>	<b>v</b>
<b>Table des matières</b>	<b>vi</b>
<b>Liste des figures</b>	<b>viii</b>
<b>Liste des tableaux</b>	<b>ix</b>
<b>Introduction</b>	<b>1</b>
<b>1 Généralités</b>	<b>3</b>
1.1 Analyse de survie . . . . .	3
1.1.1 Événement censuré et temps d'observation . . . . .	4
1.1.2 Fonction de défaillance . . . . .	6
1.1.3 Méthodes pour estimer la fonction de défaillance . . . . .	8
1.1.4 Apprentissage automatique pour l'analyse de survie . . . . .	13
1.1.5 Les métriques d'évaluation . . . . .	15
1.2 Changements de régimes et méthodes de détections . . . . .	18
1.2.1 Les changements de régimes . . . . .	18
1.2.2 Changement de régime dans la variation de la moyenne . . . . .	19
1.2.3 Changement de régime dans la variation de la variance . . . . .	20
1.3 Réseaux de neurones . . . . .	21

## TABLE DES MATIÈRES

1.3.1	Notion de neurone . . . . .	22
1.3.2	Réseau de neurones profonds . . . . .	24
1.3.3	Apprentissage des paramètres du réseau de neurones . . . . .	24
<b>2</b>	<b>Méthodologie du modèle de survie</b>	<b>28</b>
2.1	Notion d'échec de modèle pré-entraîné . . . . .	28
2.2	Nouvelle approche de prédiction d'échec de modèles prédictifs . . . . .	30
2.2.1	Entraînement des modèles prédictifs . . . . .	30
2.2.2	Génération du processus de survie . . . . .	31
2.2.3	Extraction des caractéristiques . . . . .	36
2.3	Entraînement du modèle de survie . . . . .	38
2.4	Prédiction d'échec et changements de régimes . . . . .	41
2.5	Travaux connexes et contributions . . . . .	42
<b>3</b>	<b>Résultats et discussion</b>	<b>44</b>
3.1	Données et modèles de séries temporelles . . . . .	44
3.1.1	Les modèles de séries chronologiques . . . . .	45
3.2	Résultats . . . . .	49
3.2.1	Évaluation du modèle de survie . . . . .	50
	<b>Conclusion</b>	<b>56</b>



# Liste des figures

1.1	Événements observés et d'objets censurés . . . . .	5
1.2	Contraintes pour le calcul de la métrique C-index . . . . .	16
1.3	Un neurone avec ses composantes . . . . .	22
1.4	Fonction sigmoïde $\sigma(x)$ . . . . .	23
1.5	Réseau de neurones : deux couches cachées, une d'entrée et de sortie .	25
2.1	Différentes étapes du modèle de survie . . . . .	29
2.2	Méthode de détermination du processus de survie . . . . .	33
3.1	Ensemble de séries chronologiques . . . . .	45
3.2	Segmentation des données . . . . .	49
3.3	Probabilité de survie avec <b>ARIMA</b> , <b>GARCH</b> et <b>RW</b> . . . . .	51
3.4	Probabilité de survie avec <b>TBATS</b> , <b>PROPHET</b> et <b>NN</b> . . . . .	52

# Liste des tableaux

3.1	Valeurs de la métrique CBS . . . . .	53
3.2	Valeurs de la métrique MAE . . . . .	54
3.3	Valeurs de la métrique C-index . . . . .	54

# Introduction

Grâce à l’expansion des données, nous assistons aujourd’hui plus que jamais à l’anticipation des décisions qui sont basées sur les résultats des modèles prédictifs. Cette prise de décisions en avance est rapidement adoptée par les acteurs de la finance [45, 26, 59, 67], de l’économie [113, 76, 87], de la médecine [46, 100, 88, 109], etc. Cependant, les échecs engendrés par les prédictions faites avec les modèles mathématiques sont inévitables. Ainsi, on peut se demander quelles seront les conséquences lorsque les modèles prédictifs seront en échec, c’est-à-dire ne performant plus comme voulus. Il est indéniable que l’échec d’un modèle prédictif pourrait conduire à de mauvaises décisions capables d’entraîner des conséquences néfastes. L’échec des modèles prédictifs dans ces domaines est souvent une conséquence de changements de régimes dans le marché [14], par exemple dans le cas de la finance [112]. Ainsi, être en mesure d’anticiper les échecs des modèles pré-entraînés et par conséquent les changements de régimes qui se manifestent en finance et en économie est devenu une préoccupation primordiale dans les marchés financiers [10]. Cependant, il est très difficile, voire même impossible d’anticiper ces changements de régimes. Néanmoins, nous pouvons essayer de prédire la durée de survie des modèles prédictifs utilisés pour prédire le comportement des marchés financiers.

Ce mémoire a pour but de proposer une solution d’anticipation sur la durée de survie des modèles prédictifs. Ainsi, nous avons développé un *modèle de survie* prédisant la probabilité de survie d’un modèle prédictif et ceci, en tout temps  $t$ . En prédisant les probabilités de survie, notre modèle aide, d’une autre manière, à anticiper les changements de régime pouvant être observés dans les marchés financiers. Avant d’entrer dans les détails du développement du nouveau modèle de survie, au chapitre 1, nous allons discuter d’abord d’une manière plus générale de certaines no-

## INTRODUCTION

tions requises pour la compréhension de ce modèle de survie. Ce sont les notions d'analyse de survie, les changements de régimes, les réseaux de neurones profonds et les métriques d'évaluation spécifiques aux problèmes d'analyse de survie. Au chapitre 2, nous allons discuter de la méthodologie du nouveau modèle de survie. Les résultats obtenus seront présentés au chapitre 3 et pour enfin conclure.

# Chapitre 1

## Généralités

Ce chapitre décrit, d’une manière plus générale, les notions d’analyse de survie, de changements de régimes et de réseaux de neurones. La description de ces notions permettra de faire une revue des méthodes qui existent en analyse de survie. Cette description est aussi nécessaire pour mieux comprendre les résultats obtenus avec notre modèle de survie.

### 1.1 Analyse de survie

L’analyse de survie est une branche de la science statistique dont le but est d’analyser et de modéliser le temps auquel un événement peut être observé [73]. Les modèles d’analyse de survie cherchent à estimer au bout de combien de temps et avec quelle probabilité on devrait s’attendre à observer la réalisation d’un événement donné [49]. La complexité des modèles d’analyse de survie advient lorsque les données d’entraînement présentent des individus censurés [50]. En effet, dans le contexte d’analyse de survie, l’un des défis majeurs est comment apprendre les paramètres des modèles d’analyse de survie en présence de données censurées. Un autre défis est comment estimer le temps auquel un événement peut être observé en tenant compte des individus censurés. Ce dernier constitue un problème fondamental en analyse de survie et sera l’une des cibles de notre modèle de survie. Dans la sous-section suivante, nous allons

## 1.1. ANALYSE DE SURVIE

nous intéresser aux événements censurés et aux temps de réalisation des événements observés.

### 1.1.1 Événement censuré et temps d'observation

Dans l'étude d'un problème d'analyse de survie, il arrive souvent que certains événements ne sont pas observés durant l'intervalle de temps d'étude. Ce phénomène est connu sous le nom d'une censure [56]. Trois raisons peuvent mener à la censure d'un événement : (i) lorsqu'aucun événement de l'objet en question n'a pas été observé dans l'intervalle d'étude  $T$ , encore appelé temps d'analyse de survie. Dans ce cas, on parle de *censure à droite* [69] car le temps d'observation, s'il y aura lieu, sera à droite du temps d'analyse de survie. Un exemple de ce type de censure est lorsque la durée de survie d'un patient dépasse notre temps d'analyse de survie. (ii) Il peut arriver qu'avant même d'atteindre le temps d'analyse de survie, aucun événement ne soit observé, et que nous perdions toute information sur le patient. Nous parlerons ici de *censure à gauche* [79] car le temps pendant le quel nous avons perdu d'information sur le patient se trouve à gauche du temps d'analyse de survie. Enfin, (iii) il existe une censure dite *censure par intervalle* qui se manifeste lorsque l'événement apparaît seulement dans un intervalle de temps dont les bornes sont respectivement différents des bornes de l'intervalle d'étude. Parmi ces trois types de censures d'événement, les censures à gauche restent de loin les plus fréquentes dans la vie réelle.

Il est important de noter que quelque soit le type de censure, le temps où un événement sera observé est inconnu en avance. La figure 1.1 montre les différentes censures que nous pouvons rencontrer dans l'évolution d'un ensemble d'objets. L'axe des ordonnées représente les objets à étudier, et l'axe des abscisses est le temps d'observation des événements. Seuls les objets dont les trajectoires terminent par une étoile présentent des événements observés, et le reste sont des objets censurés. Cependant, il est crucial de signaler que tout objet dont le temps d'observation correspond au temps d'étude sera aussi considéré comme un objet censuré à gauche. Cette remarque jouera un rôle fondamental dans notre modèle de survie. En effet, on verra que pour ce modèle de survie, tous les événements des modèles, qui correspondent ici aux objets, sont observés sauf ceux dont le temps d'observation coïncide avec le temps d'analyse

## 1.1. ANALYSE DE SURVIE

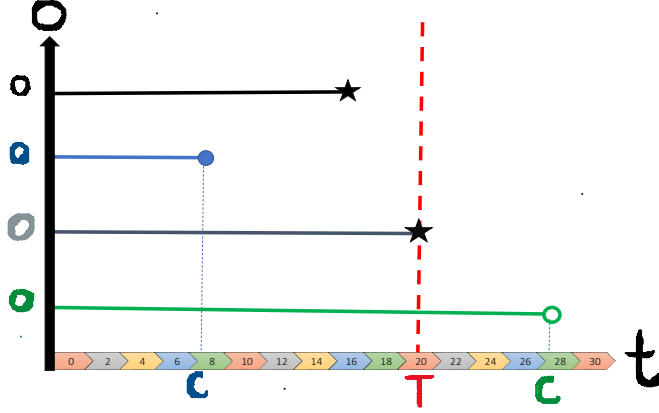


Figure 1.1 – Exemple d'événements observés et d'individus censurés dans un intervalle d'étude de vingt jours. L'axe des ordonnées représente les objets à étudier, et l'axe des abscisses est le temps d'observation des événements. Les objets, dont l'évolution se termine au moment  $t = C$  sont des objets censurés et  $t = T$  est l'intervalle d'étude représenté par la ligne rouge en pointillée. Selon que  $C > t$  ou  $C < t$ , on parlera respectivement de *censure à droite* et de *censure à gauche*. Les objets dont la trajectoire se termine pas une étoile, sont les objets dont les événements sont observées.

de survie. Compte tenu des différents types de censures, le temps d'observation des événements, qui sera noté par  $K \neq (C, T)$ , est connu seulement pour ceux observés dans l'intervalle d'étude. Pour les objets, dont le temps d'observation est plus grand que le temps d'étude ou dont on a perdu leur trace avant le temps d'observation, nous pouvons seulement avoir le temps de censure noté  $C$ , c'est-à-dire, le temps où l'événement a été censuré (ou simplement le temps où l'individu a été perdu).

D'une manière plus explicite, soit  $i$  un individu ou un objet dans un ensemble de données. Nous désignerons par  $K_i$  le temps où un événement de l'individu  $i$  est observé, et par  $C_i$ , le temps où l'individu est censuré. Les données sont qualifiées de données *censure à droite* si  $t_i = \min(K_i, C_i)$ , où  $t_i$  est le temps d'observation de l'événement ou le temps de censure de l'individu  $i$ . Il est important de noter que  $K$  et  $C$  sont des variables aléatoires. Mathématiquement, le temps d'observation de l'événement ou de censure de l'individu  $i$  est donné par :

$$t_i = \begin{cases} K_i & \text{si } \delta_i = 1, \\ C_i & \text{si } \delta_i = 0, \end{cases} \quad (1.1)$$

## 1.1. ANALYSE DE SURVIE

où  $\delta_i$  est un indicateur, c'est-à-dire, il prend respectivement les valeurs 0 ou 1 pour un individu  $i$  censuré ou dont l'événement est observé. Le but principal de l'analyse de survie est d'estimer le temps d'observation de l'événement  $K_i$  d'un individu  $i$ . Une méthode probabiliste très efficace permettant d'estimer ce temps d'observation est de calculer la fonction de défaillance. Dans la sous-section suivante, nous allons définir et étudier les propriétés de la fonction de défaillance. Cependant nous commencerons d'abord par introduire les deux fonctions qui jouent un rôle important dans l'analyse de survie : la fonction de survie et sa fonction de distribution de probabilité.

### 1.1.2 Fonction de défaillance

La fonction de défaillance se déduit de deux fonctions importantes à savoir la fonction de survie et sa fonction de distribution de probabilité. La fonction de survie [49, 107, 56, 69], que nous noterons par  $\mathcal{S}(t)$ , représente la probabilité  $\mathcal{P}(K > t)$ , que le temps d'observation d'un événement soit supérieur à une valeur  $t$  de la variable aléatoire  $K$ , c'est-à-dire :

$$\mathcal{S}(t) = \mathcal{P}(K > t). \quad (1.2)$$

La fonction de survie a les propriétés suivantes :

- (i) c'est une fonction qui décroît progressivement lorsque le temps  $t$  augmente ;
- (ii) au temps  $t = 0$ ,  $\mathcal{S}(t = 0) = 1$ , car au moment initial de l'étude, on doit supposer que tous les individus ont 100% de chance de survivre. On s'intéresse généralement à la fonction de densité de probabilité complémentaire. On l'obtient en calculant la probabilité suivante :

$$\mathcal{F}(t) = \mathcal{P}(K \leq t), \quad (1.3)$$

c'est-à-dire, le temps d'observation est inférieur au temps  $t$ . Comme l'une des propriétés de la probabilité indique que :

$$\mathcal{P}(K > t) + \mathcal{P}(K \leq t) = 1, \quad (1.4)$$

nous pouvons récrire  $\mathcal{F}(t)$  en fonction de la fonction de survie  $\mathcal{S}(t)$  comme suit :

$$\mathcal{F}(t) = 1 - \mathcal{S}(t). \quad (1.5)$$



## 1.1. ANALYSE DE SURVIE

La fonction de densité  $f(t)$  de la distribution  $\mathcal{F}(t)$  s'obtient facilement en dérivant cette dernière, soit :

$$f(t) = \begin{cases} \frac{d}{dt}\mathcal{F}(t) & \text{dans le cas continu,} \\ \lim_{\Delta t \rightarrow 0} \frac{\mathcal{F}(t+\Delta t) - \mathcal{F}(t)}{\Delta t} & \text{dans le cas discret,} \end{cases} \quad (1.6)$$

où  $\Delta t$  est un paramètre de discrétisation qui tend vers 0 dans le cas continu.

En analyse de survie, on utilise très souvent une autre fonction appelée fonction de défaillance et notée par  $h(t)$ . Elle indique en quelque sorte le taux d'échec conditionnel [98, 38, 29]. Cette fonction ne donne pas la probabilité qu'un événement apparaisse à un instant  $t$  mais représente plutôt la vraisemblance qu'un événement survive en ce moment  $t$  étant donné l'absence d'événement avant  $t$ . Nous pouvons l'écrire sous la forme :

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathcal{P}(t \leq K < t + \Delta t \mid K \geq t)}{\Delta t}. \quad (1.7)$$

En utilisant l'équation (1.3), on peut écrire la relation suivante :

$$\lim_{\Delta t \rightarrow 0} \frac{\mathcal{P}(t \leq K < t + \Delta t \mid K \geq t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{\mathcal{F}(t + \Delta t) - \mathcal{F}(t)}{\Delta t \cdot \mathcal{S}(t)} = \frac{f(t)}{\mathcal{S}(t)}. \quad (1.8)$$

Et, en se servant de l'équation (1.6), on en déduit que :

$$h(t) = \frac{f(t)}{\mathcal{S}(t)}. \quad (1.9)$$

Étant une vraisemblance,  $h(t)$  est évidemment une fonction positive. On peut aussi l'exprimer en fonction de la fonction de survie  $\mathcal{S}(t)$  en constatant que  $f(t) = -\frac{d}{dt}\mathcal{S}(t)$ . En effet, on peut facilement déduire que :

$$h(t) = -\frac{1}{\mathcal{S}(t)} \frac{d}{dt}\mathcal{S}(t) = -\frac{d}{dt} [\ln \mathcal{S}(t)]. \quad (1.10)$$

Si nous définissons la fonction cumulative de la distribution de défaillance  $h(t)$  par  $H(t) = \int_0^t h(\tau) d\tau$  [69], la fonction de survie devient :

$$\mathcal{S}(t) = e^{-H(t)}. \quad (1.11)$$

## 1.1. ANALYSE DE SURVIE

Comme la fonction cumulative a des valeurs positives, on retrouve bien la décroissance exponentielle de la fonction de survie  $\mathcal{S}(t)$  en fonction du temps. Le but de l'analyse de survie revient alors à estimer la fonction cumulative de la distribution de défaillance. Dans la sections suivante, on discutera des différentes méthodes bien connues pour trouver la distribution de  $h(t)$ .

### 1.1.3 Méthodes pour estimer la fonction de défaillance

Plusieurs méthodes ont été développées dans le but de mieux estimer la distribution de défaillance ou encore la fonction cumulative issue de cette distribution, notamment les méthodes dites non-paramétriques, semi-paramétriques, paramétriques, et plus récemment les méthodes d'apprentissage automatiques.

#### Méthodes non-paramétriques

Les méthodes non-paramétriques les plus remarquables sont la méthode de Kaplan-Meier [55, 95, 86], l'estimateur de Nelson-Aalen [80, 2, 81] et la méthode de la table de vie [15, 32, 16]. Elles sont généralement utilisées dans le cas où il y a un manque d'information quant à la distribution que devrait suivre les données. En effet, ces méthodes définissent une relation empirique pour estimer la fonction de survie  $\mathcal{S}(t)$ .

Dans le cas du Kaplan-Meier, la fonction de survie est donnée par le produit de la même fonction d'estimation. Nous pouvons l'exprimer comme suit :

$$\mathcal{S}(t) = \prod_{T_i \leq t} \left(1 - \frac{\alpha_i}{\beta_i}\right), \quad (1.12)$$

où  $T_i$ ,  $\alpha_i$  et  $\beta_i$  sont respectivement le temps d'étude à l'état  $i$ , le nombre d'événements à cet état et le nombre d'individus en risque juste avant  $T_i$ .

L'estimateur de Nelson-Aalen donne une estimation de la fonction cumulative de la distribution de défaillance sous la forme :

$$H(t) = \sum_{\tau_i \leq t} \frac{\alpha_i}{\beta_i}, \quad (1.13)$$

où  $\alpha_i$  et  $\beta_i$  sont les mêmes que dans l'équation (1.12). Juste par sa définition, on

## 1.1. ANALYSE DE SURVIE

comprend bien que l'estimateur de Nelson-Aalen est une fonction en escalier continue avec une augmentation au temps d'échec (temps d'observation). Il est important de noter que la fonction de survie est liée à la distribution de défaillance par le biais de l'équation (1.11) ( $\mathcal{S}(t) = e^{-H(t)}$ .)

Enfin, la table de vie ou encore table de mortalité, très simple, se compose d'un ensemble de lignes où chaque ligne représente une période discrète, disons une année, et des colonnes caractérisant les différents statuts de mortalité pour la période. C'est en effet, une application de la méthode de Kaplan-Meier aux données de survie groupées par intervalles. Dans la suite, nous explorerons les méthodes semi-paramétriques.

### Méthodes semi-paramétriques

Dans la catégorie des méthodes semi-paramétriques, le modèle de Cox [19, 20, 104, 4, 51] est l'approche d'analyse de régression la plus couramment utilisée. Dans l'approche de Cox, la fonction de défaillance d'un individu  $i$ , ( $i = 1, \dots, N$ ), représenté par le triplé  $(\mathbf{X}_i, t_i, \delta_i)$ , prend la forme :

$$h(t, \mathbf{X}_i) = h_0(t) \exp(\beta^T \cdot \mathbf{X}_i), \quad (1.14)$$

où  $\mathbf{X}_i = (x_{i1}, \dots, x_{ip})$  est le vecteur des variables et  $(t_i, \delta_i)$  sont définis précédemment dans l'équation (1.1);  $h_0(t)$  est la fonction de défaillance de base et  $\beta = (\beta_1, \dots, \beta_p)$ , un vecteur dont les entrées sont les paramètres du modèle. Compte tenu de l'équation (1.10), la fonction de survie s'exprime comme :

$$\mathcal{S}(t, \mathbf{X}_i) = e^{-\int_0^t h(\tau, \mathbf{X}_i) d\tau} = (e^{-\int_0^t h_0(\tau) d\tau})^{\exp(\beta^T \cdot \mathbf{X}_i)} = (e^{-H_0(t)})^{\exp(\beta^T \cdot \mathbf{X}_i)}, \quad (1.15)$$

où  $H_0(t)$  est la fonction cumulative de la distribution de défaillance de base  $h_0(t)$ . La fonction de survie devient :

$$\mathcal{S}(t, \mathbf{X}_i) = (S_0(t))^{\exp(\beta^T \cdot \mathbf{X}_i)}, \quad (1.16)$$

avec  $S_0(t) = \exp(-H_0(t))$ , la fonction de survie de base. Dans la majorité des études, l'estimateur de Breslow [68] est utilisé pour estimer  $H_0(t)$ . Dans la méthode par

### 1.1. ANALYSE DE SURVIE

estimateur de Breslow,  $H_0(t)$  est écrit comme une somme des fonctions de défaillance sous la forme :

$$H_0(t) = \sum_{T_i \leq t} \hat{h}_0(T_i, \mathbf{X}_i), \quad (1.17)$$

où  $\hat{h}_0(T_i, \mathbf{X}_i)$  est donnée par :

$$\hat{h}_0(T_i, \mathbf{X}_i) = \begin{cases} \frac{1}{\sum_{j \in R_i} \exp(\beta^T \cdot \mathbf{X}_i)} & \text{si } T_i \text{ est un temps où un événement est observé,} \\ 0 & \text{sinon.} \end{cases} \quad (1.18)$$

Dans l'équation (1.18),  $R_i$  est l'ensemble des individus qui sont en risque au temps  $T_i$ , c'est-à-dire, les individus dont un événement pourrait être observé au temps  $T_i$ . Dans ce qui suit, nous discuterons des méthodes permettant d'apprendre les paramètres  $\beta$ , qui figurent dans le modèle de Cox.

On constate que dans l'équation (1.14), la fonction de défaillance de base  $h_0(t)$  n'est pas spécifiée. Ainsi, il devient impossible d'approcher les paramètres de ce modèle à l'aide de la fonction de vraisemblance standard.  $h_0(t)$  est effectivement une fonction de nuisance tandis que les coefficients  $\beta$  sont les paramètres d'intérêts du modèle. Dans le but d'apprendre ces paramètres  $\beta$ , on peut utiliser la vraisemblance partielle proposée par Cox [19, 20]. En effet, cette vraisemblance partielle dépend seulement des paramètres  $\beta$  et non de la fonction de nuisance inconnue. Étant donné que  $h(t, \mathbf{X}_i)$  représente la probabilité qu'un individu subit un échec au temps  $t$ , la condition que cet individu survive pendant un laps de temps  $dt$ , peut s'écrire comme  $h(t, \mathbf{X}_i) dt$ ,  $dt \rightarrow 0$ . La probabilité conditionnelle d'un individu  $j$ , dont l'événement est observé au moment  $T_j$ , peut être formulée comme :

$$\frac{h(T_j, \mathbf{X}_j) dt}{\sum_{i \rightarrow R_j} h(T_j, \mathbf{X}_i) dt}. \quad (1.19)$$

La vraisemblance partielle de Cox est un produit sur les individus  $j = 1, \dots, N$  :

$$L(\beta) = \prod_{j=1}^N \left( \frac{\exp(\beta^T \cdot \mathbf{X}_j)}{\sum_{i \rightarrow R_j} \exp(\beta^T \cdot \mathbf{X}_i)} \right)^{\delta_j}. \quad (1.20)$$

### 1.1. ANALYSE DE SURVIE

Les valeurs estimées  $\hat{\beta}$  des paramètres  $\beta$  sont ceux qui minimisent le logarithme de la vraisemblance  $(-\log L(\beta))$ , soit :

$$\hat{\beta} = -\min_{\beta} \sum_{j=1}^N \delta_j \left( \mathbf{X}_j \beta - \log \sum_{i \rightarrow R_j} \exp(\mathbf{X}_i \beta) \right). \quad (1.21)$$

Il est connu que dans le cas où le nombre de variables est trop élevé par rapport au nombre d'individus, la généralisation de ce modèle de prédiction devient très difficile à cause d'un problème de sur-apprentissage. Ainsi, pour solutionner ce problème, on ajoute un terme de régularisation des paramètres  $\beta$  dans la vraisemblance. Il existe deux méthodes de régularisation dont la méthode de Lasso-Cox [106, 36] et la méthode de Ridge-Cox [40, 111, 110]. La première est une régularisation de norme  $l_1$  permettant une sélection de caractéristiques ainsi que l'estimation des paramètres  $\beta$ . La régularisation de Ridge-Cox est une régularisation de norme  $l_2$  permettant de sélectionner les caractéristiques corrélées et réduit en plus leur valeurs. D'une manière plus générale, la famille de norme  $\lambda$ , appliquée aux paramètres  $\beta$ , est donnée par :

$$\|\beta\|_{\lambda} = \left( \sum_{i=1}^p \|\beta_i\|^{\lambda} \right)^{\frac{1}{\lambda}}, \quad \lambda > 0. \quad (1.22)$$

Il devient important de préciser que les performances de ces estimateurs pénalisés dépendent considérablement de l'hyperparamètre  $\lambda$ . Sa valeur peut être estimée en faisant une validation croisée.

La version du modèle de Cox discutée plus haut est seulement applicable lorsque les variables sont indépendantes du temps. Dans le cas où ces variables changent en fonction du temps, une autre approche plus adaptée a été mise à jour. Elle est fondée sur une sorte de processus de comptage [54]. Soit :

$$\mathbf{X}(t) = (\mathbf{X}_1(t), \mathbf{X}_2(t), \dots, \mathbf{X}_{p_1}(t), \mathbf{X}'_1, \mathbf{X}'_2, \dots, \mathbf{X}'_{p_2}), \quad (1.23)$$

l'ensemble des variables avec  $p_1$  et  $p_2$  respectivement le nombres de variables dépendantes et indépendantes du temps ( $\mathbf{X}'$  représentent les variables qui sont indépendantes du temps). Le modèle de Cox de l'équation (1.14) peut être dynamisé en le

## 1.1. ANALYSE DE SURVIE

reformulant comme suit :

$$h(t, \mathbf{X}(t)) = h_0(t) \exp \left( \sum_{j=1}^{p_1} \beta_j^{(1)} \mathbf{X}_j(t) + \sum_{i=1}^{p_2} \beta_i^{(2)} \mathbf{X}'_i \right), \quad (1.24)$$

où  $\beta_j^{(1)}$  et  $\beta_i^{(2)}$  sont les coefficients correspondant respectivement à la  $j^{\text{ème}}$  et  $i^{\text{ème}}$  variable dépendante et indépendante du temps. Cependant, on s'intéressera plus souvent au taux de défaillance comme mentionné précédemment. Pour deux variables  $\mathbf{X}(t)$  et  $\mathbf{X}^*(t)$ , le taux de défaillance est donné par :

$$R(t) = \frac{h(t, \mathbf{X}^*(t))}{h(t, \mathbf{X}(t))} = \exp \left( \sum_{j=1}^{p_1} \beta_j^{(1)} [\mathbf{X}_j^*(t) - \mathbf{X}_j(t)] + \sum_{i=1}^{p_2} \beta_i^{(2)} [\mathbf{X}'_i^* - \mathbf{X}'_i] \right). \quad (1.25)$$

On précise que dans le but de trouver les paramètres optimaux  $\hat{\beta}^{(1)}$  et  $\hat{\beta}^{(2)}$ , la fonction de vraisemblance pour le modèle de Cox dépendant du temps peut être construite de la même manière et optimisée avec la même complexité temporelle que le modèle de Cox indépendant du temps. Les méthodes non-paramétriques et semi-paramétriques sont nécessaires lorsque la distribution des individus est inconnue. Dans le cas contraire, il est plus intéressant d'utiliser les méthodes paramétriques que nous discuterons dans la sous-section suivante.

### Méthodes paramétriques

Les méthodes paramétriques reposent sur le fait que les données d'entraînement suivent une distribution théorique connue [69]. Notons tout de suite que ceci constitue une limitation majeure dans l'utilisation de ces méthodes car la plus part des données rencontrées ne suivent pas une distribution particulière, ou pouvant être facilement inférée. Cependant, si la distribution est bien connue, ces modèles offrent des alternatives plus efficaces que les modèles de Cox non-paramétriques et semi-paramétriques [73, 47]. Les distributions les plus utilisées sont la distribution normale [28], la distribution exponentielle et la distribution logistique [62, 18]. On peut trouver aussi la distribution Weibull et la distribution log-logistique [28]. Dans les méthodes paramétriques, grâce à la connaissance de la distribution, on peut facilement déduire la vraisemblance associée à cette distribution. Pour trouver les paramètres optimaux, on

## 1.1. ANALYSE DE SURVIE

maximise ainsi cette vraisemblance. D’une manière générale, la vraisemblance peut s’écrire sous la forme [70] :

$$L(\beta) = \prod_{\delta_i=1} f(K_i, \beta) \prod_{\delta_i=0} S(K_i, \beta), \quad (1.26)$$

dans l’équation (1.26),  $\prod_{\delta_i=1} f(K_i, \beta)$  et  $\prod_{\delta_i=0} S(K_i, \beta)$  sont respectivement les probabilités jointes pour les individus observés et censurés. Une fois que les paramètres optimaux,  $\beta = \beta^*$ , trouvés, la fonction de défaillance  $h(t)$  pourra être estimée en utilisant les équations (1.16) et (1.17). Il est important de noter que l’approche des méthodes paramétriques est aussi connue sous le nom de l’approche de Cox. Dans la sous-section suivante, nous allons introduire l’utilisation de l’apprentissage automatique dans la prédiction incluant des données censurées. L’apprentissage automatique via les réseaux de neurones sera discuté largement dans la section 1.3.

### 1.1.4 Utilisation de l’apprentissage automatique pour l’analyse de survie

La capacité de l’apprentissage automatique à modéliser les relations non linéaires a suscité beaucoup d’intérêt dans plusieurs domaines de recherche notamment dans l’analyse de survie [114, 21, 72, 61, 5]. La difficulté majeure de l’apprentissage automatique dans les problèmes d’analyse de survie est le traitement des individus censurés. En effet, dans beaucoup de problèmes d’analyse de survie, il n’y a pas toujours un nombre raisonnable d’instances dans chaque dimension alors que ceci constitue une caractéristique fondamentale pour un bon fonctionnement de l’apprentissage automatique [115]. Dans ce qui suit, nous allons présenter les méthodes d’apprentissage automatique les plus utilisées dans le domaine de l’analyse de survie.

#### Arbres de survie

D’une manière plus générale, les méthodes d’arbres partitionnent les données selon un critère bien défini. Ainsi, les données semblables ou ayant les mêmes propriétés sont regroupées dans un même noeud selon ce critère. La différence entre les méthodes usuelles d’arbres, c’est-à-dire, faisant l’usage aux données sans censures et

## 1.1. ANALYSE DE SURVIE

celles d'analyse de survie [65, 42, 23, 6], se situe dans le choix du critère de partitionnement [33, 12, 97]. En effet, même si la partition fondée sur une valeur seuil permet de classer les données usuelles, elle n'est pas construite pour pouvoir tenir en compte ni l'interaction entre les caractéristiques, ni le fait que les données sont censurées. Les critères de partitionnement dans les arbres développés pour l'analyse de survie peuvent être groupés en deux catégories : (i) ceux qui minimisent une fonction de perte utilisant l'homogénéité intra-noeuds [33, 27, 64] et (ii) ceux qui maximisent l'hétérogénéité inter-noeuds en se basant sur des tests statistiques [22, 13]. Une autre caractéristique arbitraire des arbres d'analyse de survie est dans la sélection du dernier noeud. Ici, certaines procédures très connues en réseaux de neurones telles que le *forward* ou le *backward* servent à l'optimisation de l'arbre de survie [6].

### Méthodes bayésiennes

Les deux méthodes bayésiennes les plus utilisées aujourd'hui sont le modèle de Bayes naïf et le réseau bayésien. La première méthode, très populaire aussi en apprentissage automatique, voit sa croissance en médecine [11] et dans la prédiction des temps de survie [3]. Ces deux méthodes fournissent en général un lien entre la probabilité a posteriori et la probabilité a priori qui est en effet l'un des principes fondamentaux dans la théorie bayésienne.

### Machines à vecteurs de supports

La méthode des machines à vecteurs de support (SVM) est déjà utilisée surtout dans la classification mais aussi dans les problèmes de régression [101, 103]. Cependant, dans les problèmes d'analyse de survie, le fait que les individus sont censurés est omise par les méthodes standards de machines à vecteurs de supports [94]. Une version adaptée de la méthode SVM consiste à ajouter une contrainte supplémentaire sur la fonction de perte. Ceci, afin de maintenir l'ordre requis entre deux individus comparables. Cependant, dans la situation où la dimension des données est considérable, cette contrainte additionnelle constitue un problème de ressource [44]. Une modification plus appropriée prend avantage de la version standard de la SVM mais avec une fonction de perte asymétrique [60]. Cette formulation de la fonction de perte



## 1.1. ANALYSE DE SURVIE

permet de tenir en compte aussi bien les instances censurées que celles observées.

Les réseaux de neurones artificiels sont aussi utilisés pour modéliser des problèmes d’analyse de survie. Nous discuterons largement des réseaux de neurones profonds dans la Section 1.3. Dans la sous-section suivante, nous allons discuter certaines métriques permettant de tester la performance des modèles d’analyse de survie, c’est-à-dire, comment ces modèles se généralisent lorsque appliqués sur de nouvelles données.

### 1.1.5 Les métriques d’évaluation

Dans les modèles d’analyse de survie, la formulation standard des métriques d’évaluation doit changer due à la présence des individus censurés. Ainsi, certaines métriques spécifiques ont été définies telles que l’*indice de concordance*, le *score de Brier*, et l’*erreur moyenne absolue* [114]. Les deux dernières ont été adaptées pour tenir compte de l’information sur les individus censurés.

#### L’indice de concordance

L’indice de concordance ou encore la probabilité de concordance, notée par C-index, mesure le risque relatif d’un événement qui survient pour des individus différents [39, 37]. En considérant deux individus, leur temps de survie peut être ordonné selon que les deux individus sont tous observés ou que le temps  $K_i$ , où les événements des deux individus sont observés, est inférieur au temps de censure  $C_i$ . À la figure 1.2, on montre ces deux différents cas de situation ; en (a) où toutes les instances sont observées et en (b) où deux instances sont censurées, indiquées par des points en couleur rouge. Dans cette figure, les courbes en bleu indiquent les différentes comparaisons possibles entre les individus qui entrent dans le calcul du C-index [96]. Il devient important de noter que dans le calcul du C-index, certaines restrictions devront être imposées au niveau de la comparaison entre deux individus (voir la figure 1.2). Pour deux observations  $y_1(t)$  et  $y_2(t)$ , au moment  $t$ , et leur prédiction correspondante  $\hat{y}_1(t)$  et  $\hat{y}_2(t)$  respectivement, on calcule le C-index en évaluant la probabilité suivante :

$$\text{C-index} = \mathcal{P}(\hat{y}_1(t) > \hat{y}_2(t) \mid y_1(t) \geq y_2(t)). \quad (1.27)$$

## 1.1. ANALYSE DE SURVIE

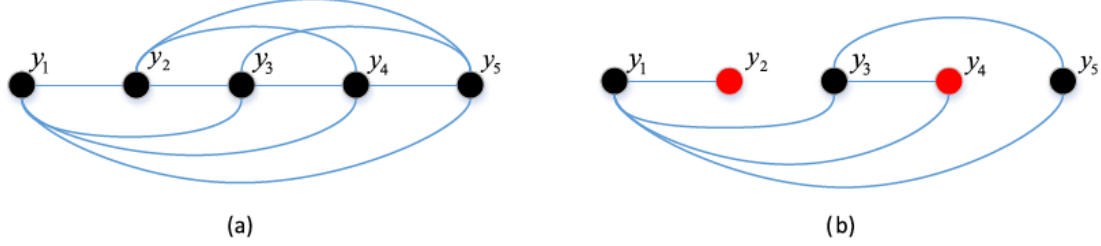


Figure 1.2 – Les contraintes fixées lors du calcul de la métrique C-index. En (a), dans le cas où aucun individu n’est censuré et où deux instances sont censurées en (b) représentées par les points rouges. On suppose que les événements sont de sorte que  $y_1 < y_2 < y_3 < y_4 < y_5$ . Figure adaptée de l’article [114].

Notons que dans un cadre plus général, l’équation (1.27) correspond exactement à la fonction d’efficacité du récepteur, plus fréquemment désignée sous le terme de *courbe de ROC*. Cependant, dans le cas des méthodes d’analyse de survie, le C-index prend la forme :

$$\text{C-index} = \frac{1}{\tilde{N}} \sum_{i:\delta_i=1} \sum_{j:y_i(t) < y_j(t)} \mathbb{I}(\mathcal{P}(\hat{y}_j(t) | X_j) > \mathcal{P}(\hat{y}_i(t) | X_i)), \quad (1.28)$$

où  $i, j \in \{1, \dots, N\}$ ,  $N$  est le nombre d’individus,  $y_i$  est l’observation de l’individu  $i$ ,  $\mathcal{P}(\hat{y}_i(t) | X_i)$  est la probabilité de survie étant donné la variable  $X_i$ ,  $\tilde{N}$  est le nombre de paires comparables et  $\mathbb{I}$  désigne la fonction d’indicateur.  $\sum_{i:\delta_i=1}$  est une somme sur tous les individus  $i$ , dont les événements sont observés ( $\delta_i = 1$ ), et  $\sum_{j:y_i(t) < y_j(t)}$  est une somme sur les individus  $j$  tel que, l’observation  $y_i(t)$  soit inférieure à l’observation  $y_j(t)$  au temps  $t$ . Par sa définition, on comprend que cette métrique peut être seulement utilisée dans les modèles de survie dont la sortie est une distribution de probabilité. Dans le cas de notre modèle de survie, ceci est exactement le cas et le C-index nous servira à mesurer sa généralisation.

### Le score de Brier

Comme dans le cas de la métrique C-index, le score de Brier [9] est seulement efficace lorsque nous aimerions évaluer un modèle dont la sortie finale est une distribution de probabilités. Si on dénote par  $\hat{y}_i(t)$  la prédiction au temps  $t$  faite par

## 1.1. ANALYSE DE SURVIE

un modèle de survie sur une nouvelle instance  $i$  de variable  $X_i, i = 1, \dots, N$  et  $y_i(t)$  l'observation correspondante, on définit le score de Brier  $BS'(t)$  par :

$$BS'(t) = \frac{1}{N} \sum_{i=1}^N [\hat{y}_i(t) - y_i(t)]^2, \quad (1.29)$$

où  $y_i(t) \in \{0, 1\}$ ,  $\hat{y}_i(t) \in [0, 1]$ , et  $N$  est le nombre d'individus. Le score de Brier a été adapté pour évaluer les méthodes d'analyse de survie en introduisant un poids  $w_i(t)$  pour chaque instance  $i$  au temps  $t$ . Dans ce cas, on peut définir un autre score de Brier  $BS(t)$  sous la forme :

$$BS(t) = \frac{1}{N} \sum_{i=1}^N w_i(t) [\hat{y}_i(t) - y_i(t)]^2. \quad (1.30)$$

Les poids sont estimés en utilisant la distribution de Kaplan-Meier  $KM(\alpha_i)$  décrite dans la section 1.1.3. Pour un individu  $i$  représenté par l'ensemble  $(\mathbf{X}_i, \alpha_i, \delta_i)$ , les poids sont estimés sous la forme :

$$w_i(t) = \begin{cases} \frac{\delta_i}{KM(\alpha_i)} & \text{si } \alpha_i \leq t, \\ \frac{1}{KM(\alpha_i)} & \text{si } \alpha_i > t. \end{cases} \quad (1.31)$$

Dans l'équation (1.31), les individus censurés ( $\delta_i = 0$ ) ont une contribution nulle mais sont indirectement utilisés dans la détermination de la distribution de Kaplan-Meier  $KM(\alpha_i)$ .

### Moyenne de l'erreur absolue

La moyenne de l'erreur absolue (MAE) est définie comme la moyenne de la différence en valeur absolue entre la valeur prédite et l'observation. Elle exclut les instances censurées et s'exprime comme suit :

$$MAE = \frac{1}{N} \sum_{i=1}^N \delta_i |\hat{y}_i(t) - y_i(t)|. \quad (1.32)$$

Dans l'équation (1.32),  $\delta_i$  est donné dans les sous-sections précédentes et  $\hat{y}_i(t)$  et  $y_i(t)$  sont respectivement la valeur prédite et la cible au temps  $t$  de l'individu  $i$ .

## 1.2. CHANGEMENTS DE RÉGIMES ET MÉTHODES DE DÉTECTIONS

Après avoir discuté de l'analyse de survie incluant les modèles non-paramétriques, semi-paramétriques et paramétriques ainsi que les métriques permettant d'évaluer la généralisation de ces modèles de survie, nous allons discuter des changements de régimes dans la section suivante. En effet, le modèle de survie, comme déjà mentionné au début, est développé pour prédire la durée de survie des modèles pré-entraînés. Ce résultat peut être utilisé dans le but de prédire les changements de régimes. Ainsi, il devient fondamental de se familiariser avec ces changements de régimes.

## 1.2 Changements de régimes et méthodes de détections

L'un des défis majeurs de modéliser les séries temporelles est surtout lorsqu'il y a la présence de changements de régimes. On cherche souvent à détecter d'abord ces changements de régimes dans le but de mieux comprendre les caractéristiques des séries pour développer ainsi un modèle qui sera capable de se généraliser en présence de ces changements de régimes. Cependant, même l'information qu'un modèle ait concernant la présence des changements de régimes lors de son entraînement ne permet pas toujours à ce modèle de mieux se généraliser. La meilleure solution consiste à anticiper tout changement de régime. Dans ce qui suit, nous allons d'abord donner une introduction aux changements de régimes et une description simplifiée des modèles de détection de changements de régimes. Comment prédire ces changements de régimes sera l'un des objections du chapitre 2.

### 1.2.1 Les changements de régimes

Récemment, le concept de changement de régime est devenu populaire dans l'interprétation des variations longues dans les données de séries temporelles. Un changement de régime intervient dans une série temporelle lorsque le système passe d'un état  $s_t$  au temps  $t$  à un état  $s_{t+k}$  au temps  $t + k$  avec  $k \geq 1$ . Les propriétés des deux états doivent être différentes, ce qui permet même de différencier les régimes. Théoriquement, ce concept est mal défini. Cependant, dans le monde de la finance, des changements de régimes sont parfaitement observables dans le marché. Elles cor-

## 1.2. CHANGEMENTS DE RÉGIMES ET MÉTHODES DE DÉTECTIONS

respondent en fait à des changements brusques du marché et ont généralement pour conséquences des pertes financières.

Un certain nombre de méthodes ont été développées pour détecter les changements de régime ou les points de changement dans les séries chronologiques. Une grande majorité de ces méthodes sont conçues pour trouver des changements dans la moyenne et la variance. Ces deux méthodes ont un inconvénient commun : leurs performances se détériorent considérablement si les points de changement de régime sont trop proches de la fin des données de séries temporelles. Une solution possible est de traiter les observations en séquentielle c'est-à-dire les observations sont testées en fonction du temps pour déterminer la validité de l'hypothèse nulle  $H_0$  correspondant à l'existence d'un changement de régime. Il y a trois résultats possibles du test : accepter  $H_0$ , rejeter  $H_0$  ou continuer le test.

### 1.2.2 Détection de changement de régime dans la variation de la moyenne

Pour cette méthode, le test est basé sur le test  $t$ -Student. Soit  $X_T = x_1, x_2, \dots, x_T$  une série temporelle avec les valeurs  $x_i, (i = 1, 2, \dots, T)$ , ajoutées régulièrement à la série  $X_T$ . Si nous ajoutons une nouvelle donnée, on procède à une vérification pour déterminer si elle représente un écart statistiquement significatif par rapport à la valeur moyenne,  $\bar{x}_a$ , du régime actuel. Selon le test  $t$ -Student, la différence,  $\mathcal{D}$ , entre le régime actuel et ce nouveau régime, de valeur moyenne  $\bar{x}_b$ , doit satisfaire la condition [91] :

$$\mathcal{D} = |\bar{x}_a - \bar{x}_b| = t\sqrt{2\bar{s}_l^2/l}, \quad (1.33)$$

pour être statistiquement significative. Dans cette équation,  $t$  est la valeur de la distribution  $t$ -Student de dimension  $2l-2$ , et de probabilité de confiance  $p$ . On suppose ici que les deux régimes ont la même variance  $\bar{s}_l$ , calculée dans un intervalle  $l < T$  de la série  $X_T$ . Nous pouvons vérifier que  $\mathcal{D}$  est toujours une constante. À l'instant actuel  $t_a$ , la valeur moyenne du nouveau régime est inconnue, mais on sait qu'elle devrait être supérieure à un niveau critique  $\bar{x}_c^+$ , si le décalage est vers le haut, ou, inférieure à  $\bar{x}_c^-$ , si le décalage est vers le bas. Les niveaux critiques  $\bar{x}_c^+$  et  $\bar{x}_c^-$  sont données par les

## 1.2. CHANGEMENTS DE RÉGIMES ET MÉTHODES DE DÉTECTIONS

relations [91] :

$$\begin{aligned}\bar{x}_c^+ &= \bar{x}_a + \mathcal{D}, \\ \bar{x}_c^- &= \bar{x}_a - \mathcal{D}.\end{aligned}\tag{1.34}$$

Si la valeur actuelle  $x_a$  est supérieure à  $\bar{x}_c^+$  ou inférieure à  $\bar{x}_c^-$ , le temps  $t_a$  sera marqué comme un point de changement potentiel, et les données suivantes sont utilisées pour rejeter ou accepter cette hypothèse. Le test consiste à calculer un indice de changement de régime (ICR), qui représente une somme cumulée d'anomalies, normalisée par rapport au niveau critique  $\bar{x}_c$ . Cet indice est donné par [91] :

$$\text{ICR} = \frac{1}{l\bar{s}_l} \sum_{i=t_a}^m (x_i - \bar{x}_c), \quad m = t_a, t_a + 1, \dots, t_a + l - 1, \tag{1.35}$$

où  $l$  est négatif, si  $\bar{x}_c = \bar{x}_c^+$ , et positif, si  $\bar{x}_c = \bar{x}_c^-$ . Si entre  $t_a$  à  $t_a + l$ , l'hypothèse nulle sur l'existence d'un décalage par rapport à la moyenne, au moment  $t_a$ , est rejetée, la valeur  $x_a$  est incluse dans le régime actuel. Dans le cas contraire, le temps  $t_a$  est déclaré comme un point de changement de régime.

### 1.2.3 Détection de changement de régime dans la variation de la variance

La procédure de détection de changement de régime dans la variance est similaire à celle utilisée avec la moyenne de la série. La différence réside dans le type de test utilisé. Ici, nous utiliserons le F-test. Dans cette méthode, on suppose que la valeur moyenne de la série chronologique est nulle, c'est-à-dire, les décalages par rapport à la moyenne sont supprimés de la série chronologique d'origine. Le F-test consiste à comparer le rapport des variances,  $r$ , des deux régimes [91] :

$$r = \frac{\bar{s}_a^2}{\bar{s}_b^2}, \tag{1.36}$$

où  $\bar{s}_i$  est la variance au moment  $t_i$ . Nous avons  $r > r_c$  ou  $r < r_c$ , où,  $r_c$  est une valeur critique de la distribution du F-test. À l'instant  $t_a$ , la variance  $\bar{s}_b$  est inconnue. Pour

### 1.3. RÉSEAUX DE NEURONES

que le nouveau régime soit statistiquement différent du régime actuel, la variance  $\bar{s}_b$  doit être supérieure à une variance critique  $\bar{s}_c^+$ , si la variance est croissante, ou inférieure à  $\bar{s}_c^-$ , si la variance est décroissante. Les variances critiques sont données par [91] :

$$\bar{s}_c^+ = \bar{s}_a^2 r_c, \quad (1.37)$$

$$\bar{s}_c^- = \bar{s}_a^2 r_c. \quad (1.38)$$

Si au moment  $t_a$ , la valeur actuelle  $x_a$  satisfait à l'une des conditions,  $x_a^2 > \bar{s}_c^{2+}$  ou  $x_a^2 < \bar{s}_c^{2-}$ , elle est marquée comme un point de décalage potentiel, et les valeurs suivantes  $x_{a+1}, x_{a+2}, \dots$  sont utilisés pour vérifier cette hypothèse. La vérification est basée sur la somme résiduelle de l'indice des carrés (RSSI), définie par :

$$\text{RSSI} = \frac{1}{l} \sum_{i=t_a}^m (x_i^2 - \bar{s}_c^2), \quad m = t_a, t_a + 1, \dots, t_a + l - 1. \quad (1.39)$$

La règle de décision est similaire à celle des décalages de la moyenne décrite dans la section précédente.

## 1.3 Réseaux de neurones

Les réseaux de neurones sont une partie d'un champ de recherche plus large qu'on appelle l'intelligence artificielle [82, 90, 75, 48]. Ils ont été intuitivement formulés à partir de l'activité cérébrale humaine. En effet, cette dernière peut être modélisée sur deux niveaux purement abstraits. D'une part, un modèle issu d'un neuroscientifique tiendrait en compte la dynamique mais surtout les propriétés biologiques détaillées de l'activité cérébrale. D'un autre côté, un scientifique cognitif modéliserait l'activité cérébrale sans faire référence aux composantes biologiques. Entre ces deux modèles, vient le réseau de neurones artificiels [89, 63, 93, 31]. Un neurone biologique reçoit plusieurs signaux à travers les synapses en contact avec ses dendrites et envoie un seul flux de potentiel d'action à travers son axone. Ainsi, on peut comprendre qu'un neurone est une fonction permettant de recevoir une information complexe pour ensuite la traduire en une information plus simple et décisive. Dans ce qui suit, nous

### 1.3. RÉSEAUX DE NEURONES

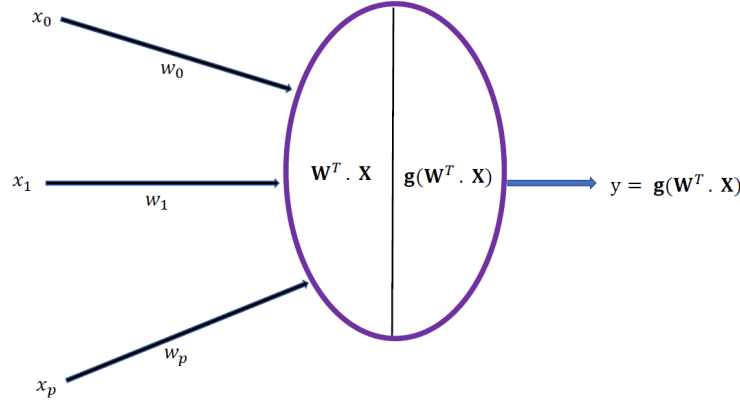


Figure 1.3 – Un neurone avec un vecteur d’entrée  $\mathbf{X} = (x_0, x_1, \dots, x_P)$ , un vecteur de poids  $\mathbf{W} = (w_0, w_1, \dots, w_P)$ , et une fonction d’activation  $g$ . La sortie du neurone  $y = g(\mathbf{W}^T \cdot \mathbf{X})$  est un scalaire qui servira d’entrée pour les autres neurones des couches suivantes.

nous intéresserons à la notion de neurone et de réseau de neurones profonds.

#### 1.3.1 Notion de neurone

Le neurone est l’une des parties fondamentales et élémentaires d’un réseau de neurones. Il constitue la partie qui mimique le neurone biologique constitué d’une entrée ou le signal d’entrée, les poids synaptiques, modélisant l’intensité du signal et d’une fonction d’activation permettant d’activer ou désactiver le neurone. La figure 1.3 donne un exemple de neurone avec ses poids synaptiques, sa fonction d’activation et son signal de sortie. On caractérise un neurone par ses composantes  $(\mathbf{X}, \mathbf{W}, g, y)$  où  $\mathbf{X} = (x_0, x_1, \dots, x_P)$  est le vecteur d’entrée,  $\mathbf{W} = (w_0, w_1, \dots, w_P)$  est le vecteur des poids synaptiques, et  $g$  est la fonction d’activation. La sortie du neurone  $y$  donnée par :

$$y = g(\mathbf{W}^T \cdot \mathbf{X}) = g\left(\sum_{i=1}^P w_i x_i\right), \quad (1.40)$$

est un scalaire qui servira d’entrée pour les autres neurones des couches suivantes. Dans l’équation (1.42), la fonction d’activation  $g$  joue un rôle important dans les réseaux de neurones. Elle permet non seulement à activer ou à désactiver le neurone, mais sert surtout à traiter les relations non linéaires entre les entrées. Il existe plusieurs



### 1.3. RÉSEAUX DE NEURONES

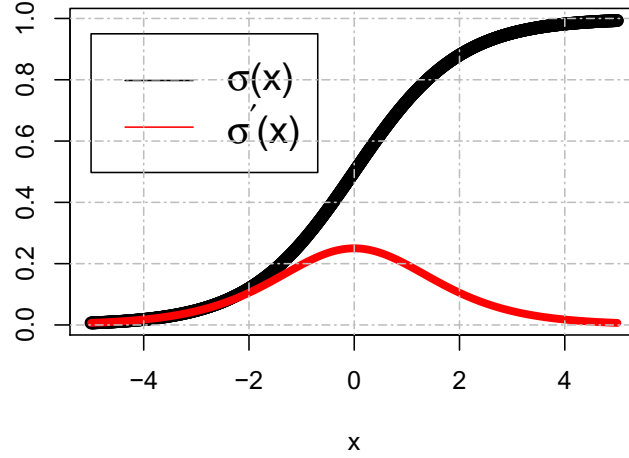


Figure 1.4 – La fonction sigmoïde  $\sigma(x)$  en noire et sa première dérivée en rouge.

fonctions d'activation. Cependant, la fonction sigmoïde définie par :

$$\sigma(\mathbf{W}^T \cdot \mathbf{X}) = \frac{1}{1 + e^{-\mathbf{W}^T \cdot \mathbf{X}}}, \quad (1.41)$$

dont nous portons notre attention dans ce mémoire, est largement utilisée dans les réseaux de neurones. À la figure 1.4, on montre le comportement de cette fonction sigmoïde et sa dérivée. Dans le cas particulier où on utilise la fonction sigmoïde comme fonction d'activation, on parlera de neurone sigmoïde. Il est défini comme une combinaison linéaire entre le vecteur d'entrée  $\mathbf{X}$  et le vecteur de des poids  $\mathbf{W}$  suivie par l'application de la fonction d'activation  $\sigma(\mathbf{W}^T \cdot \mathbf{X})$  comme décrite par l'équation (1.41). Les valeurs de sortie d'un neurone sigmoïde s'apparentent à des probabilités appartenant donc à l'intervalle  $[0, 1]$ . Les sorties plus proches que 1 ont plus de chance d'être observées que celles plus proche que 0. L'avantage fondamental de la fonction sigmoïde est sa continuité en fonction des poids. En effet, on verra que cette propriété joue un rôle crucial dans l'apprentissage des paramètres  $\mathbf{W}$  du modèle. Cependant, il est important de noter que la dérivée devient presque nulle lorsque ses paramètres tendent vers  $\pm\infty$ . Ce comportement devra être pris en compte

### 1.3. RÉSEAUX DE NEURONES

lors de l'apprentissage des paramètres du réseau de neurones en adoptant surtout un processus de mise en échelle. Dans la suite, nous étudierons le branchement de plusieurs neurones conduisant aux réseaux de neurones profonds.

#### 1.3.2 Réseau de neurones profonds

Un réseau de neurones est une structure formée de plusieurs neurones rangés sous forme de couches. Un réseau de neurones profonds comporte des couches cachées en plus de la couche d'entrée et de sortie à la différence d'un réseau de neurone multicouches. Ce dernier peut être composé seulement de la couche d'entrée et de sortie. Un exemple de réseau de neurones est présenté à la figure 1.5. Il est composé d'une couche d'entrée de  $P$  neurones, d'une première et deuxième couche cachée de  $N$  et  $M$  neurones respectivement et d'une couche de sortie de  $K$  neurones.  $\mathbf{W}_1$ ,  $\mathbf{W}_2$ ,  $\mathbf{W}_3$  sont les matrices des poids entre les couches. Les valeurs de ces matrices de poids peuvent être positives ou négatives et encouragent respectivement une activation ou une inhibition des neurones. Le comportement de la fonction d'activation  $g$  définit certains propriétés du réseau de neurones. En effet, le réseau aura un comportement linéaire ou non-linéaire respectivement selon que la fonction d'activation est linéaire ou non-linéaire. Le modèle linéaire est obtenu lorsque la sortie du neurone est sous la forme :

$$y = \sum_{i=1}^P w_i x_i, \quad (1.42)$$

correspondant, par exemple, à une fonction d'activation identité. Ainsi, on voit bien que le réseau de neurones est en fait une généralisation du modèle linéaire capable de traiter les relations non-linéaire via sa fonction d'activation  $g(x)$ .

#### 1.3.3 Apprentissage des paramètres du réseau de neurones : algorithme de rétropropagation

L'apprentissage des paramètres est le processus permettant de trouver les meilleurs valeurs des poids  $\mathbf{W}$  où nous avons aussi inclus le biais dans cette matrice de paramètres. Dans la suite,  $c = 0, 1, \dots, C$  désigne respectivement la couche d'entrée, la

### 1.3. RÉSEAUX DE NEURONES

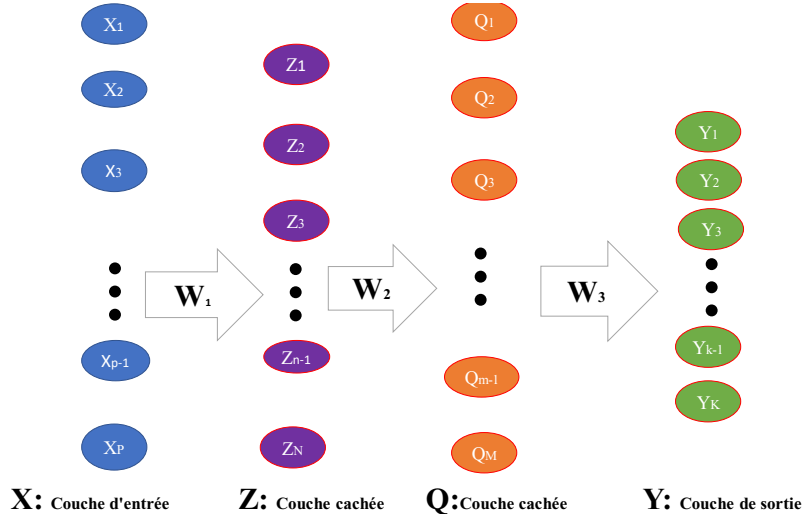


Figure 1.5 – Réseau de neurones composé d'une couche d'entrée de  $P$  neurones, de deux couches cachées de  $N$  et  $M$  neurones respectivement, et d'une couche de sortie de  $K$  neurones. Les matrices des poids sont respectivement  $\mathbf{W}_1$ ,  $\mathbf{W}_2$ , et  $\mathbf{W}_3$ , de la couche d'entrée à la couche de sortie. Partant de la représentation d'un neurone, figure 1.3, nous représentons les liens individuels entre les couches par une seule flèche.

première couche cachée ainsi de suite jusqu'à la couche de sortie où  $c = C$ . On notera par  $d^{(c)}$  le nombre de neurones de la couche  $c$ ,  $w_{ij}^{(c)}$  le poids associé au  $i^{\text{ème}}$  neurone de la couche  $c - 1$  au  $j^{\text{ème}}$  neurone de la couche  $c$ , avec  $0 \leq i \leq d^{(c-1)}$  et  $1 \leq j \leq d^{(c)}$ .  $x_j^{(c)}$  représentera la sortie du  $j^{\text{ème}}$  neurone de la couche  $c$ . Ainsi, à la figure 1.5,  $x_j^{(c)}$  correspondent à  $X_1, X_2, \dots, X_P, Z_1, Z_2, \dots, Z_N, Q_1, Q_2, \dots, Q_M$  et  $Y_1, Y_2, \dots, Y_K$  respectivement pour la couche d'entrée, la première couche cachée, la deuxième couche cachée et la couche de sortie. Considérons le  $j^{\text{ème}}$  neurone de la couche  $c$ . La valeur de ce neurone est une combinaison des poids, des valeurs des neurones précédents, et de l'application de la fonction d'activation. Elle s'écrit sous la forme :

$$x_j^{(c)} = g\left(s_j^{(c)}\right) \quad \text{avec} \quad s_j^{(c)} = \sum_{i=1}^{d^{(c-1)}} w_{ij}^{(c)} x_i^{(c-1)}. \quad (1.43)$$

Pour trouver les valeurs  $x_j^{(c)}$  dans l'équation (1.43), et en particulier  $x_1^{(C)}, x_2^{(C)}, \dots, x_{d^{(C)}}^{(C)}$  où  $d^{(C)} = K$  à la figure 1.5, on fait une propagation avant dans le réseau. La propagation avant est la première étape dans l'apprentissage des paramètres du modèle.

### 1.3. RÉSEAUX DE NEURONES

Elle permet de prédire une valeur de la cible et donc à préparer la deuxième étape, c'est-à-dire, la propagation rétro. Dans la propagation rétro, on repasse dans le réseau tout en calculant le gradient de la fonction dite de perte. Ce repassage dans le réseau permet de mettre à jour les paramètres du réseau. Soit  $L(\mathbf{W})$  cette fonction de perte mesurant une certaine proximité entre la cible et la sortie du réseau  $Y = x^{(C)}$ . Une méthode de minimisation, dite la méthode de la descente de gradient, passe par le calcul du gradient de la fonction de perte, soit  $\nabla_{\mathbf{W}} L(\mathbf{W})$ . Une dérivation en chaîne permet de calculer les dérivées de cette fonction de perte par rapport aux poids  $w_{ij}^{(c)}$ , connectant le  $i^{\text{ème}}$  neurone de la couche  $c - 1$  au  $j^{\text{ème}}$  neurone de la couche  $c$  :

$$\frac{\partial L(\mathbf{W})}{\partial w_{ij}^{(c)}} = \frac{\partial L(\mathbf{W})}{\partial s_j^{(c)}} \frac{\partial s_j^{(c)}}{w_{ij}^{(c)}}. \quad (1.44)$$

Par le biais de l'équation (1.43), on déduit que :

$$\frac{\partial s_j^{(c)}}{w_{ij}^{(c)}} = x_i^{(c-1)}. \quad (1.45)$$

En plus, si nous notons par  $\delta_j^{(c)} = \frac{\partial L(\mathbf{W})}{\partial s_j^{(c)}}$ , la sensibilité de l'erreur par rapport au signal  $s_j^{(c)}$ , l'équation (1.44) devient :

$$\frac{\partial L(w)}{\partial w_{ij}^{(c)}} = \delta_j^{(c)} x_i^{(c-1)}. \quad (1.46)$$

L'évaluation de  $\nabla_w L(w)$  serait complète si les  $\delta_j^{(c)}$  étaient connus. Ils peuvent être obtenus en appliquant un algorithme dit *rétropropagation* qui commence par déterminer  $\delta_j^{(C)}$  de la couche de sortie. La détermination de  $\delta_j^{(c)}$  dépend de la forme de la fonction de perte. Si nous supposons une fonction de perte quadratique de la forme :

$$L(w) = \frac{1}{2} \sum_{j=1}^{d^{(C)}} \left( x_j^{(C)} - \hat{y}_j \right)^2, \quad (1.47)$$

### 1.3. RÉSEAUX DE NEURONES

où  $\hat{y}_j = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_{d^{(c)}})$  est la cible correspondant à l'entrée  $x_j$ , et  $x_j^{(C)} = g(s_j^{(C)})$ , nous obtenons :

$$\delta_j^{(C)} = (x_j^{(C)} - \hat{y}_j) \frac{dg(s_j^{(C)})}{ds_j^{(C)}}. \quad (1.48)$$

Dans le cas où la fonction d'activation  $g$  est la fonction sigmoïde  $\sigma(w)$  de l'équation (1.41), l'équation (1.48) va s'écrire sous la forme :

$$\delta_j^{(C)} = (x_j^{(C)} - \hat{y}_j) \sigma(s_j^{(C)}) (1 - \sigma(s_j^{(C)})). \quad (1.49)$$

Une fois que  $\delta_j^{(C)}$  de la couche de sortie est connue, on cherchera ensuite à exprimer  $\delta_j^{(c-1)}$  de la couche  $c-1$  en fonction de  $\delta_j^{(c)}$  de la couche  $c$  en partant de la définition :

$$\delta_j^{(c-1)} = \frac{\partial L(w)}{\partial s_i^{(c-1)}} = \sum_{j=1}^{d^{(c)}} \frac{\partial L(w)}{\partial s_j^{(c)}} \frac{\partial s_j^{(c)}}{\partial s_i^{(c-1)}} = \sum_{j=1}^{d^{(c)}} \delta_j^{(c)} \frac{\partial s_j^{(c)}}{\partial s_i^{(c-1)}}. \quad (1.50)$$

L'inconnu dans l'équation (1.50) est la sensibilité du signal de la couche  $c$  par rapport au signal de la couche précédente  $c-1$ . En dérivant l'équation (1.43) sous la forme :

$$\frac{\partial s_j^{(c)}}{\partial s_i^{(c-1)}} = \frac{\partial}{\partial s_i^{(c-1)}} \left( \sum_{i=1}^{d^{(c-1)}} w_{ij}^{(c)} g(s_i^{(c-1)}) \right) = w_{ij}^{(c)} \frac{dg(s_j^{(c-1)})}{ds_j^{(c-1)}}, \quad (1.51)$$

et si la fonction d'activation est la fonction sigmoïde, l'équation (1.50) prend la forme :

$$\delta_i^{(c-1)} = \sigma(s_i^{(c-1)}) (1 - \sigma(s_i^{(c-1)})) \sum_{j=1}^{d^{(c)}} \delta_j^{(c)} w_{ij}^{(c)}. \quad (1.52)$$

Partant des valeurs initiales des poids et d'un taux d'apprentissage  $\eta > 0$ , les poids optimaux du réseau de neurones entraînés avec l'algorithme de la descente de gradient sont obtenus récursivement par la relation (pour l'itération  $(n+1)$ ) :

$$w_{ij}^{(c)}(n+1) = w_{ij}^{(c)}(n) - \eta \delta_j^{(c)}(n) x_i^{(c-1)}(n). \quad (1.53)$$

Les hyperparamètres comme  $\eta$  et le nombre de couches cachées  $C$  sont généralement obtenus en faisant une validation croisée.

# Chapitre 2

## Méthodologie du modèle de survie

Le modèle de survie est le principal résultat de ce mémoire. Nous allons discuter ici les différentes étapes qui ont conduit à son développement. Cependant, nous nous intéresserons d'abord à la notion d'échec de modèles pré-entraînés.

### 2.1 Notion d'échec de modèle pré-entraîné

Au cours des dernières décennies, plusieurs modèles ont été inventés pour améliorer leur capacité de prédiction [71, 24, 77, 57]. Certains modèles sont plus pratiques dans la mesure où ils peuvent réduire ce qu'on appelle la *probabilité d'échec* en s'appuyant sur plusieurs paramétrisations possibles. La stratégie la plus courante à ce jour, et qui s'est avérée efficace, consiste à utiliser un modèle d'ensemble [84, 74, 58, 99], c'est-à-dire, en combinant plusieurs modèles de prédiction. Malgré leur succès, on sait peu de choses sur les raisons de leurs échecs, et si ces derniers peuvent être anticipés. Les échecs de modèle se produisent lorsque le modèle déjà entraîné est incapable de faire de bonnes prédictions après un moment  $t$ . Divers phénomènes peuvent être à l'origine d'un tel échec. Les principales raisons sont :

- (i) soit le modèle n'est pas correctement entraîné,
- (ii) soit le modèle n'est pas adapté à un cas particulier d'un problème ou,
- (iii) soit il y a une différence importante entre les propriétés statistiques des don-

## 2.1. NOTION D'ÉCHEC DE MODÈLE PRÉ-ENTRAÎNÉ

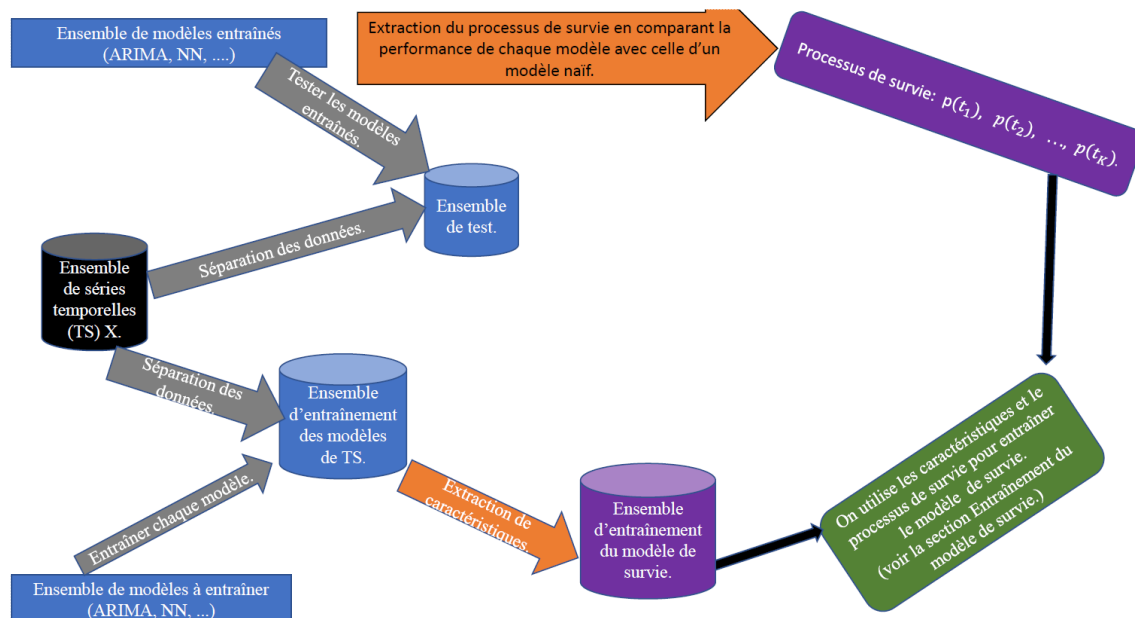


Figure 2.1 – Les différentes étapes du modèle de survie. Les modèles de séries chronologiques sont entraînés sur un ensemble d’entraînement. Leur évaluation sur un ensemble de test permet de constituer le processus de survie  $\mathcal{P}(t_1), \mathcal{P}(t_2), \dots, \mathcal{P}(t_K)$  (voir la figure 2.2). Un ensemble de caractéristiques extraites de l’ensemble d’entraînement en plus des processus de survie permettent d’entraîner et de tester notre modèle de survie.

nées d’entraînement et de test.

En effet, dans le cas des séries chronologiques, on assiste souvent à une dérive conceptuelle [83], et il faut en plus noter leurs propriétés statistiques inhabituelles, comme la non érgodicité, l’hétérostédasticité et la non stationnarité [102].

Être en mesure de prédire l’échec d’un modèle pré-entraîné a plusieurs applications surtout dans le domaine de la finance, de l’économie, et dans les situations où, une décision basée sur les prédictions d’un tel modèle, doit être prise. La détermination des causes exactes de l’échec d’un modèle donné est loin d’être triviale même si nous connaissons les principales raisons en général. Néanmoins, nous pouvons estimer dans quel contexte l’échec d’un modèle pré-entraîné a plus de probabilité d’être observé et évaluer si on est capable de les anticiper en première place. Dans ce mémoire, nous proposons un modèle d’analyse de survie pour anticiper les échecs des modèles pré-entraînés. La section suivante présentera ce modèle de survie avec les différentes

## 2.2. NOUVELLE APPROCHE DE PRÉDICTION D'ÉCHEC DE MODÈLES PRÉDICTIFS

étapes ayant conduit à son développement.

## 2.2 Nouvelle approche de prédiction d'échec de modèles prédictifs

Nous décrivons ici les différentes étapes de notre modèle de survie pour la prédiction des échecs de modèles pré-entraînés. Nous discuterons comment les échecs de modèles sont directement liés aux changements de régimes. La figure 2.1 présente, schématiquement, les différentes phases de notre modèle de survie. Dans cette figure, nous commençons par séparer les données en un ensemble d'entraînement et de test. Les modèles dont nous allons tester leur échec sont entraînés en utilisant les données d'entraînement. Les modèles déjà entraînés sont ensuite tester sur l'ensemble de test. Durant ce test, nous comparons l'erreur faite par chaque modèle avec l'erreur d'un modèle naïf. Ceci permet de construire une matrice, qu'on appelle la matrice du processus de survie de dimension  $N_{te}^u \times K$  où  $N_{te}^u$  est la dimension de l'ensemble de test et  $K$ , le temps d'étude. Les entrées de cette matrice sont des probabilités  $p(t) \in \{0, 1\}$  pour  $t = 1, 2, \dots, K$ . Nous calculons certaines caractéristiques à partir des données d'entraînement. Ces caractéristiques, ainsi que le processus de survie, sont utilisés pour entraîner notre modèle de survie. Dans la suite, nous allons présenter plus en détails les différentes étapes de notre modèle de survie.

### 2.2.1 Entraînement des modèles prédictifs

La première étape du modèle de survie est l'entraînement des modèles dont nous souhaitons prédire l'échec. Dans le test de notre modèle de survie, ces modèles seront des modèles de prédiction sur des séries temporelles. Nous discuterons de ces modèles dans la suite. Cette première étape se fait de la manière habituelle : un modèle est entraîné sur un ensemble d'entraînement de séries chronologiques. Toutes techniques comme la validation croisée peuvent être utilisées dans le but d'avoir une bonne précision du modèle lors de son évaluation sur l'ensemble de test. Dans la suite, nous désignerons par modèle pré-entraîné un modèle déjà entraîné sur l'ensemble d'entraînement.



## 2.2. NOUVELLE APPROCHE DE PRÉDICTION D'ÉCHEC DE MODÈLES PRÉDICTIFS

### 2.2.2 Génération du processus de survie

La deuxième étape de notre modèle de survie est la génération du processus de survie. Ici, un processus de survie est une séquence d'événements, de probabilités d'observation  $p(t) \in \{0, 1\}$  binaire. La valeur  $p(t) = 1$  indique qu'un événement a été observé au temps  $t$  et  $p(t) = 0$ , dans le cas contraire. En analyse de survie, ce processus de survie (différent du modèle de survie) est généralement connu car il est présent dans les données d'entraînement. En effet, le processus de survie représente les cibles pour les modèles d'analyse de survie. Cependant, pour les modèles de séries chronologiques financières, ce processus de survie n'est pas observé et donc, nous devons trouver une méthode permettant de l'inférer.

Étant donné un ensemble de séries chronologiques d'entraînement, le modèle de survie associe à chaque série  $s$  une séquence de probabilités  $p_s(t_1), p_s(t_2), \dots, p_s(t_K)$  pour un intervalle de temps  $K$ . Cette séquence de probabilités est le processus de survie et constituera ainsi les cibles du modèle de survie. Le processus de survie  $p_s(t_\beta)$ , pour  $\beta = 1, 2, \dots, K$ , indique si oui ou non un événement est observé au temps  $t_\beta$  avec la probabilité  $p_s(t_\beta)$ .  $p_s(t_\beta)$  prend la valeur 1 lorsque nous avons observé l'événement et 0 dans le cas contraire. Dans le cas de notre modèle de survie, on dira qu'un événement est observé lorsque les prédictions du modèle pré-entraîné ne sont plus correctes, c'est-à-dire, le modèle pré-entraîné a subi un échec. Pour juger le degré d'échec d'un modèle pré-entraîné, nous allons comparer l'erreur du modèle pré-entraîné avec celle d'un modèle de référence. Nous avons utilisé ici une métrique qui est la moyenne de l'erreur absolue échelonnée (MASE) pour comparer les erreurs des modèles de référence et pré-entraîné. Elle mesure le rapport entre l'erreur du modèle pré-entraîné  $e(t)$  et l'erreur du modèle de référence  $e_r(t)$  au moment  $t$ . Au temps  $t < K$ , elle s'exprime de la manière suivante :

$$\text{MASE}(t) = \frac{e(t)}{e_r(t)}. \quad (2.1)$$

où l'erreur au moment  $t$  est donnée par :

$$e(t) = | \hat{y}(t) - y(t) | \quad \text{et,} \quad e_r(t) = | \hat{y}_r(t) - y(t) |, \quad (2.2)$$

## 2.2. NOUVELLE APPROCHE DE PRÉDICTION D'ÉCHEC DE MODÈLES PRÉDICTIFS

avec  $y(t)$  et  $\hat{y}_{(r)}(t)$  sont respectivement, l'observation et la valeur prédite et  $|a|$  correspond à la valeur absolue de  $a$ . Une valeur de la  $MASE(t)$  inférieure à 1 s'obtient seulement si la valeur prédite par le modèle pré-entraîné est plus proche de la vraie valeur que celle prédite par le modèle de référence. Dans cette condition, aucun événement n'est observé et la probabilités binaire du processus de survie  $p(t) = 1$ . Cette affectation des probabilités du processus de survie continue,  $p(t+1) = 1, p(t+2) = 1, \dots$ , jusqu'à ce qu'un événement soit observé au moment  $t'$ , c'est-à-dire, la valeur de la  $MASE(t')$  est supérieure à 1, et à ce moment  $t'$ , la probabilité binaire  $p(t') = 0$ . La figure 2.2 présente une description graphique de la génération du processus de survie. Dans cette figure, un modèle de série temporelle est entraîné sur un ensemble d'entraînement composé de plusieurs séries temporelles. Ce modèle entraîné est ensuite testé sur un ensemble de test. Si l'erreur du modèle, au moment  $t$ , est plus petite que l'erreur d'un modèle de référence,  $p(t) = 1$ , sinon  $p(t) = 0$ . On répète ce processus pour tout  $t$  dans l'intervalle d'étude. La figure montre un exemple de séquence où au temps  $t' = 4$  jours, la  $MASE(t')$  est plus élevée que 1. Elle montre aussi un exemple de séquence où aucun événement n'a été observé sur tout l'intervalle de temps d'étude de sept jours. Dans ce cas  $p(t) = 1$ , pour  $t = 1, 2, \dots, 7$ , indiquant que le modèle pré-entraîné est toujours plus performant que le modèle de référence. Dans cette situation, on dira que le modèle pré-entraîné est censuré, une propriété que nous avons discuté à la section 1.1.1.

L'algorithme 1 montre comment obtenir la matrice du processus de survie dans le cas où le modèle de référence est le modèle naïf [92]. La ligne  $j$  de la matrice du processus de survie contient l'étiquette, au temps  $t$ , de la série d'entraînement  $j$  pour notre modèle de survie. Notons bien que l'erreur du modèle naïf est calculée en utilisant les données d'entraînement qui sont plus représentatives pour ce modèle. Cet algorithme indique clairement que lorsque le modèle pré-entraîné est moins performant que le modèle de référence au temps  $t' < K$ , il restera moins performant pour tout  $t$  tel que  $t' < t \leq K$ . Cependant, cette propriété n'est pas toujours vérifiée dans les séries temporelles, compte tenu de leurs propriétés souvent différentes de celles des autres type de données. En effet, le modèle pré-entraîné peut bien redevenir plus performant que le modèle de référence dans un autre intervalle de temps de début supérieur à  $t'$ , par exemple, de  $t' + 1$  à  $t' + \alpha$  pour  $\alpha < K$ . Cette situation arrive souvent lorsqu'au

## 2.2. NOUVELLE APPROCHE DE PRÉDICTION D'ÉCHEC DE MODÈLES PRÉDICTIFS

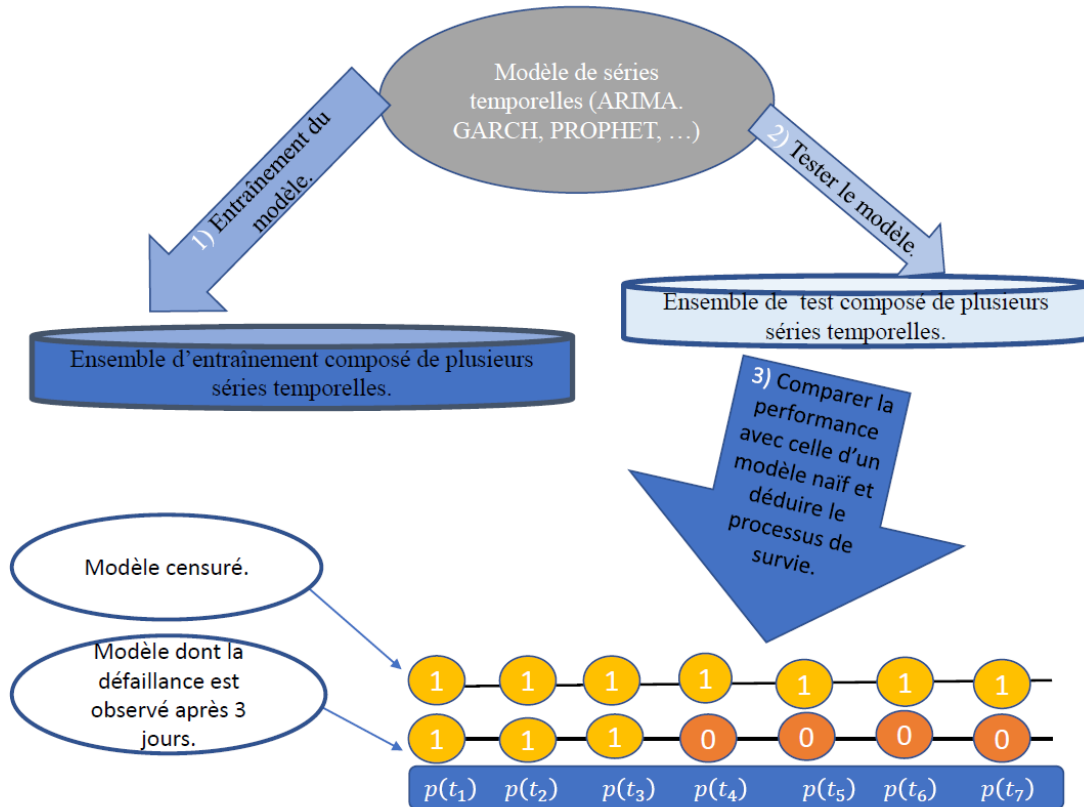


Figure 2.2 – Méthode de détermination du processus de survie. Le modèle de série chronologique ait d'abord entraîné sur l'ensemble d'entraînement. On teste ensuite le modèle pré-entraîné sur l'ensemble de test et on compare sa performance avec un modèle naïf, c'est-à-dire, on calcule la MASE par rapport au modèle naïf. Si la MASE est plus petit que 1, le processus de survie prend la valeur 1 sinon, il est égal à 0 (voir l'algorithme 1).

## 2.2. NOUVELLE APPROCHE DE PRÉDICTION D'ÉCHEC DE MODÈLES PRÉDICTIFS

temps  $t'$ , un événement particulier s'est produit durant l'enregistrement de la série temporelle.

---

**Algorithme 1 :** Algorithme pour estimer la matrice du processus de survie.

---

**Entrée :** Ensemble d'observation  $y_s(t)$ ,  $s = 1, \dots, N_{tr}^u$ ,  $N_{tr}^u$  est la dimension de l'ensemble de test, ensemble de prédiction  $\hat{y}_s(t)$  et  $\hat{y}_{r,s}(t)$ , respectivement par le modèle de prédiction et de référence.

**Sortie :** Matrice du processus de survie : Matrice PS de dimension  $N_{tr}^u \times K$ .

```

1 pour  $s = 1, \dots, N_{tr}^u$  faire
2   pour  $t = 1, \dots, K$  faire
3      $e_s(t) = |y_s(t) - \hat{y}_s(t)|$ 
4      $e_{r,s}(t) = |y_s(t) - \hat{y}_{r,s}(t)|$ 
5      $MASE(t) = \frac{e_s(t)}{e_{r,s}}$ 
6     si  $MASE(t) < 1$  alors
7        $PS[s, t] = 1$ 
8     sinon
9       pour  $t = t', \dots, K$  faire
10         $PS[s, t'] = 0$ 

```

---

## 2.2. NOUVELLE APPROCHE DE PRÉDICTION D'ÉCHEC DE MODÈLES PRÉDICTIFS

---

**Algorithme 2 :** Algorithme de raffinement de la matrice de processus de survie PS (voir algorithme 1.) La signification de  $t_\beta$  est donnée dans le texte.

---

**Entrée :** Matrice du processus de survie : Matrice PS de dimension  $N_{te}^u \times K$ ,  
un intervalle de temps  $t_\beta$ .

**Sortie :** Matrice du processus de survie raffinés : Matrice P de dimension  
 $N_{te}^u \times K$ .

```

1 pour  $s = 1, \dots, N_{te}^u$  faire
2   pour  $t = 1, \dots, K$  faire
3     pour  $t' = t + 1, \dots, t_\beta$  faire
4       si  $PS[s, t] < PS[s, t']$  alors
5          $P[s, t'] = 1$ 
6       sinon
7         pour  $t' = t, \dots, K$  faire
8            $P[s, t'] = PS[s, t]$ 

```

---

Pour tenir compte de ce cas particulier, nous devons relaxer la contrainte imposant un non retour de la meilleure performance du modèle pré-entraîné. Imaginons que nous aimerions faire une prédiction dans une fenêtre de temps  $K$  en utilisant un modèle pré-entraîné quelconque en comparant sa performance avec celle d'un modèle de référence. Si on découvre que le modèle pré-entraîné a survécu ou est plus performant que le modèle de référence au temps  $t_\beta, t_\beta + 2$  et  $t_\beta + 3$ , et non au temps  $t_\beta + 1$ , nous pouvons simplement supposer que l'échec à ce temps  $t_\beta + 1$  est accidentel dans le sens où un événement particulier devrait être responsable de l'échec du modèle pré-entraîné. Dans le but de solutionner ce problème, on procède de la manière suivante : pour une série chronologique de l'ensemble de test, si un modèle pré-entraîné ne performe pas mieux que le modèle de référence pour un intervalle de temps fixe  $\beta$ , il doit rester moins performant que le modèle de référence pour le reste du temps d'étude, c'est-à-dire aux temps  $t_\beta$  pour  $\beta < K$ . Juste dans le développement d'un modèle de survie,  $\beta$  se comporte comme un hyper-paramètre et sa valeur devra être fixée. L'algorithme 2 permet d'obtenir le processus de survie final pour une valeur de  $\beta$  donnée. Dans le

## 2.2. NOUVELLE APPROCHE DE PRÉDICTION D'ÉCHEC DE MODÈLES PRÉDICTIFS

suite, nous discuterons de l'extraction des caractéristiques permettant d'entraîner le modèle de survie.

### 2.2.3 Extraction des caractéristiques

Le modèle de survie peut être entraîné soit en utilisant les mêmes données d'entraînement ayant servi à entraîner les méthodes dites pré-entraînées. Cependant, dans certaines circonstances, il sera toujours plus efficace d'utiliser les caractéristiques extraites des séries chronologiques ayant permis à entraîner les modèles pré-entraînés. Les caractéristiques sont des variables mathématiques calculées à partir de la série pour ensuite représenter la série. Le modèle de survie accepte généralement les caractéristiques globales. Ces dernières quantifient les patrons existants dans les séries sur la longueur globale de la série contrairement aux caractéristiques locales. Une simple application des caractéristiques globales, autre que la réduction de la dimension, est lorsqu'on souhaite comparer deux séries. Ici, des séries ayant les mêmes caractéristiques globales ont tendance à présenter les mêmes propriétés. Il existe plusieurs méthodes mathématiques d'extraction de caractéristiques à partir de séries temporelles. Elles réduisent la dimension et permettent surtout de représenter un ensemble de séries chronologiques par une matrice simple et donc permet de préparer les entrées dans des méthodes de prédiction comme l'apprentissage automatique profond. Dans cette représentation, le nombre de lignes de la matrice est égal au nombre de séries temporelles, et le nombre de colonnes correspond au nombre de caractéristiques extraites. Une manière simple d'extraire les caractéristiques globales et leur signification est décrite dans le package R *tsfeatures* [41]. Dans la suite, nous nous intéressons à certaines de ces caractéristiques.

La mesure de la distribution d'une série chronologique, en utilisant la moyenne ou la variance, peut donner des informations sur certaines propriétés de la série. Si on désigne par  $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$ , la moyenne d'une série temporelle stationnaire d'ordre deux  $X = X_1, X_2, \dots, X_N$  de longueur  $N$ , on calcule la variance,  $s_X^2$ , comme suit [30] :

$$s_X^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2. \quad (2.3)$$

## 2.2. NOUVELLE APPROCHE DE PRÉDICTION D'ÉCHEC DE MODÈLES PRÉDICTIFS

La variance est indépendante de l'ordre temporel de la série. Elle informe si la série est stationnaire en restant constante et non infinie. Il existe, cependant, une métrique permettant de tester la stationnarité de la moyenne d'une série temporelle. Soit la série temporelle  $\overline{X_{1:w}}, \overline{X_{w+1:2w}}, \dots, \overline{X_{(m-1)w+1:mw}}$ , de  $m$  fenêtres d'intervalle de temps distincts, de largeur  $w$ , et de variance  $s_{X_w}^2$ . Ici  $\overline{X_{1:w}}$  est la moyenne de la série  $X_1, X_2, \dots, X_w$ . La métrique  $\text{StatAv}(\tau)$  permettant de tester la stationnarité d'ordre deux de cette série temporelle de moyenne est donnée par [30] :

$$\text{StatAv}(\tau) = \frac{s_{X_w}}{s_X}, \quad (2.4)$$

où  $\tau = mw$  est la longueur de la série temporelle.  $\text{StatAv}(\tau)$  permet de capturer comment les composantes temporelles de la série varient au fur du temps.

L'auto-corrélation,  $C(\tau)$ , est une autre caractéristique qui mesure la corrélation entre les entrées d'une série temporelle dans un intervalle de temps  $\tau$ . Elle est estimée de la manière suivante [30] :

$$C(\tau) = \frac{1}{s_X^2(N - \tau)} \sum_{t=1}^{N-\tau} (X_t - \bar{X}) (X_{t+\tau} - \bar{X}). \quad (2.5)$$

La dernière caractéristique que nous discuterons ici est l'entropie  $H(m, r)$  [85], différent de l'entropie de Shannon, donnée par :

$$H(m, r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \ln \frac{A_i}{N - m + 1} - \frac{1}{N - m} \sum_{i=1}^{N-m} \ln \frac{A_i}{N - m}. \quad (2.6)$$

où  $m$  indique la longueur d'une séquence de patrons,  $r$  est la distance de seuil entre deux patrons consécutifs,  $A_i$  est le nombre de vecteurs qui sont à une distance  $r$  de  $u_m$ . En effet, on représente ici les séries temporelles comme une intégration temporelle en termes de l'ensemble de vecteurs  $u_m$ .  $H(m, r)$  est utilisée pour quantifier combien une série chronologique peut être prédite.

Il existe d'autres caractéristiques que nous pouvons utiliser pour représenter les séries chronologiques. Le package *tsfeatures* de R donne une description plus exhaustive de ces caractéristiques. Nous noterons que notre modèle de survie peut accueillir d'autres caractéristiques endogènes ou exogènes, c'est-à-dire, respectivement, des ca-

### 2.3. ENTRAÎNEMENT DU MODÈLE DE SURVIE

caractéristiques calculées à partir de la série temporelle ou basées sur d'autres propriétés indépendantes de la série temporelle.

## 2.3 Entraînement du modèle de survie

Les trois étapes telles que l'entraînement du modèle de séries chronologiques, la génération du processus de survie et l'extraction des caractéristiques, discutées plus haut, sont les entrées de notre modèle de survie. La dernière étape est d'apprendre les paramètres de notre modèle de survie. Dans cette optique, nous utilisons ici plusieurs classificateurs de durée de survie [117]. Considérons une classification binaire effectuant la prédiction de l'échec d'un modèle pré-entraîné dans un intervalle de temps donné  $K$ . Nous voudrions en effet répondre à la question : est-ce que le modèle pré-entraîné restera sans échec sur les  $t$  prochains jours, mois, etc. ( $0 < t < K$ ) ? Nous transformons ainsi simplement le problème d'origine, de prédiction de durée de vie restante d'un modèle pré-entraîné, en une série de problèmes de classifications binaires. Ceci est vrai tant que chacun a une probabilité de sortie spécifique à une durée de vie restante, disons  $\tau$ , qui n'est pas antérieure à  $t$ . Ainsi, une fonction cumulative  $\mathcal{F}(t)$  peut être définie à partir de cette probabilité de sortie. En effet, si l'échec du modèle pré-entraîné intervient dans un intervalle de temps  $[0, t]$ , la fonction  $\mathcal{F}(t)$  sera donnée par l'équation (1.3) lorsque  $\tau = K$  :

$$\mathcal{F}(t) = \mathcal{P}(\tau < t). \quad (2.7)$$

où  $\mathcal{P}(\tau < t)$  est la probabilité que la durée de vie restante  $\tau$  du modèle pré-entraîné soit inférieure à  $t$ . Un classificateur sera considéré comme une machine qui génère plusieurs probabilités de classification binaire sur des instantanés de temps disjoints  $t_1 < t_2 < \dots < t_K$  chacune représentant la probabilité qu'un événement ne soit pas observer au temps  $t_\alpha$ ,  $\alpha = 1, 2, \dots, K$ , c'est-à-dire, qu'un modèle pré-entraîné ne subit pas un échec à  $t_\alpha$ . Les probabilités de sortie forment un vecteur de probabilités et peuvent être interprétées comme la précision du modèle pré-entraîné en fonction du temps.

Notre modèle de survie est un réseau de neurones dont la couche de sortie corres-



### 2.3. ENTRAÎNEMENT DU MODÈLE DE SURVIE

pond aux probabilités  $\mathcal{P}_s(t_\alpha)$  qu'un modèle pré-entraîné reste sans échec au temps  $t_\alpha$ ,  $s = 1, 2, \dots, N$  où  $N$  est le nombre de séries chronologiques. La figure 1.5 montre un exemple de réseau de neurones dont  $Y_1 = \mathcal{P}_s(t_1)$ ,  $Y_2 = \mathcal{P}_s(t_2), \dots, Y_K = \mathcal{P}_s(t_K)$  pour une série chronologique  $s$ . Il est important de noter que le nombre de couches cachées du modèle de survie est un hyper-paramètre qui doit être déterminé par validation croisée et donc spécifique aux données d'entrée. Aussi, il est important de noter que ces probabilités correspondent aux probabilités de sortie de notre modèle de survie et non aux probabilités  $p(t)$ , ayant servi à l'entraînement. Nous utiliserons dans ce mémoire, le réseau de neurone avec deux couches cachées comme le montre la figure 1.5. Les propriétés générales des réseaux de neurones profonds sont discutées dans la section 1.3. Dans la suite de cette section, on montrera comment le modèle de survie est entraîné, c'est-à-dire, comment apprendre les paramètres de ce réseau de neurones. Le réseau est composé d'une couche d'entrée de  $P$  neurones, d'une première et deuxième couche cachée de  $N$  et  $M$  neurones respectivement et d'une couche de sortie de  $K$  neurones. Précisons que  $N$  et  $K$  sont ceux déjà utilisés. Le rôle de la couche d'entrée est de distribuer les entrées dans la première couche cachée. Les valeurs de cette dernière sont obtenues par une somme de toutes les entrées pondérées par les vecteurs de la matrices de poids  $\mathbf{W}_1$ . Un biais  $\mathbf{b}_1$  est ajouté et une fonction d'activation (la sigmoïde  $\sigma(x)$ , voir la figure 1.4) est appliquée. La première couche cachée agit maintenant comme une entrée pour la deuxième couche cachée dont les valeurs des neurones est une combinaison des entrées de la première couche cachée pondérée par les vecteurs de la matrices de poids  $\mathbf{W}_2$ , un biais  $\mathbf{b}_2$  est ajouté et une fonction d'activation sigmoïde est appliquée. Finalement, on obtient les probabilités de survie  $\mathcal{P}_s(t_k) = Y_k$  en sommant sur toutes les entrées de la deuxième couche pondérées par les poids  $\mathbf{W}_3$  et en ajoutant un biais  $\mathbf{b}_3$  et en appliquant une fonction d'activation sigmoïde. Le choix de la fonction sigmoïde comme fonction d'activation est motivé par sa propriété probabiliste mais aussi sa composante exponentielle qui peut être utilisée dans la mesure du risque d'événement. Étant donné un ensemble de  $\mathcal{N}$  caractéristiques  $\mathbf{C} = \mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_N$  extraites d'un ensemble de séries chronologiques, la probabilité de survie d'un modèle pré-entraîné au temps  $t_\alpha$  connaissant

### 2.3. ENTRAÎNEMENT DU MODÈLE DE SURVIE

$\mathbf{C}$ ,  $(\mathbf{W}_1, \mathbf{b}_1)$ ,  $(\mathbf{W}_2, \mathbf{b}_2)$ , et  $(\mathbf{W}_3, \mathbf{b}_3)$  s'exprime comme :

$$\begin{aligned}\mathbf{G}(\mathbf{C}, \mathbf{W}, \mathbf{b}) &= \sigma(\mathbf{W}_2 \cdot \sigma(\mathbf{W}_1 \cdot \mathbf{C} + \mathbf{b}_1) + \mathbf{b}_2) \\ \mathcal{P}(t_\alpha | \mathbf{C}, \mathbf{W}, \mathbf{b}) &= \sigma(\mathbf{W}_3 \cdot \mathbf{G}(\mathbf{C}, \mathbf{W}, \mathbf{b}) + \mathbf{b}_3).\end{aligned}\tag{2.8}$$

Dans l'Éq. (2.8), la dépendance en temps des caractéristiques vient des séries temporelles dont elles sont extraites et donc le risque d'événement a aussi une dépendance temporelle. On peut noter que dans cette équation, l'estimation des probabilités de survie se fait par une propagation en avant. On utilise ici la divergence de censure de Kullback-Leibler comme définie dans [117] pour prédire des probabilités aussi proche possible que les processus de survie (voir section 2.2.2). Pour un processus de survie ou encore les cibles  $p_s(t_\alpha) \in \{0, 1\}$  et les probabilités prédites  $\mathcal{P}_s(t_\alpha | \mathbf{C}, \mathbf{W}, \mathbf{b})$  d'une série chronologique  $s$  (représentée par les caractéristiques extraites), la divergence de Kullback-Leibler censurée est donnée par :

$$\begin{aligned}d_s(t_\alpha, \mathbf{W}, \mathbf{b}) &= p_s(t_\alpha) \ln \frac{p_s(t_\alpha)}{\mathcal{P}_s(t_\alpha | \mathbf{C}, \mathbf{W}, \mathbf{b})} \\ &+ (1 - p_s(t_\alpha)) \ln \frac{1 - p_s(t_\alpha)}{1 - \mathcal{P}_s(t_\alpha | \mathbf{C}, \mathbf{W}, \mathbf{b})}.\end{aligned}\tag{2.9}$$

Dans l'équation (2.9), le premier terme est utilisé lorsque  $p_s(t_\alpha) = 1$ , et le deuxième terme lorsque  $p_s(t_\alpha) = 0$ . Nous pouvons trouver les valeurs optimales des poids, et des biais en minimisant l'erreur moyenne calculée à partir de la divergence de Kullback-Leibler comme suit :

$$E(\mathbf{W}, \mathbf{b}) = \frac{1}{N_{tr}} \sum_{s=1}^{N_{tr}} \sum_{\alpha}^K d_s(t_\alpha, \mathbf{W}, \mathbf{b}),\tag{2.10}$$

où  $N_{tr}$  est le nombre de séries chronologiques dans l'ensemble d'entraînement. Au minimum de la fonction de perte  $E(\mathbf{W}, \mathbf{b})$ , on trouve les meilleurs paramètres  $\mathbf{W}^*, \mathbf{b}^*$  de  $\mathbf{W}, \mathbf{b}$  du modèle de survie telles que les distributions  $\mathcal{P}_s(t_\alpha | \mathbf{C}, \mathbf{W}, \mathbf{b})$  et  $p_s(t_\alpha)$  soient aussi proches que possible.

Nous optimisons le réseau de neurones en utilisant l'algorithme d'Adam [52], c'est un algorithme pour l'optimisation du gradient de premier ordre basé sur des esti-

## 2.4. PRÉDICTION D'ÉCHEC ET CHANGEMENTS DE RÉGIMES

mations adaptatives des moments d'ordre inférieur. Il a l'avantage d'être simple à mettre en œuvre et ne nécessite pas beaucoup de ressources et convient donc pour des problèmes importants en termes de données et/ou de paramètres. Dans l'algorithme d'Adam, au pas de temps  $\tau$ , le gradient :

$$g_\tau = \min_{\mathbf{W}, \mathbf{b}} E(\mathbf{W}, \mathbf{b}), \quad (2.11)$$

est déterminé et les estimations de premier et deuxième moments biaisés sont respectivement mises à jour selon :

$$\begin{aligned} m_t &\leftarrow (\lambda_1 \cdot m_{t-1} + (1 - \lambda_1) \cdot g_\tau) / (1 - \lambda_1^\tau) \\ v_t &\leftarrow (\lambda_2 \cdot v_{t-1} + (1 - \lambda_2) \cdot g_\tau^2) / (1 - \lambda_2^\tau), \end{aligned} \quad (2.12)$$

pour les hyper-paramètres  $\lambda_1$  et  $\lambda_2$  et  $\lambda^\tau$  est une puissance. Pour l'itération  $\tau$ , les paramètres  $\mathbf{W}, \mathbf{b}$  du modèle sont mises à jour sous la forme :

$$(\mathbf{W}_\tau, \mathbf{b}_\tau) \leftarrow (\mathbf{W}_{\tau-1}, \mathbf{b}_{\tau-1}) - \tilde{\alpha} \cdot \frac{m_\tau}{\sqrt{v_\tau} + \epsilon}, \quad (2.13)$$

où  $\tilde{\alpha}$  est le pas en temps et  $\epsilon$  une valeur infinitésimale pour éviter la divergence du rapport. Dans la section suivante, nous allons montrer comment la prédiction de l'échec d'un modèle pré-entraîné est liée à la prédiction des changements de régimes dans les séries chronologiques.

## 2.4 Relation entre Prédiction d'échec de modèle et changements de régimes

Supposons qu'un modèle souhaitable pour un problème donné soit bien entraîné sur un ensemble de données d'entraînement, c'est-à-dire la précision du modèle sur un ensemble de données de test est acceptable ou encore le modèle est plus précis qu'un modèle de référence. Entraîner un modèle correspond à la recherche des meilleurs paramètres du modèle où l'erreur entre les valeurs prédites et réelles est la plus petite possible. De toute évidence, lorsqu'un modèle pré-entraîné n'est plus capable de faire

## 2.5. TRAVAUX CONNEXES ET CONTRIBUTIONS

des prédictions précises après un temps  $t_\alpha$  sur un même ensemble de données de test, ceci devrait être dû à des paramètres inappropriés en ce temps  $t_\alpha$  du modèle déjà entraîné (pré-entraîné). La question que nous pouvons se demander est la suivante : comment les paramètres d'un modèle pré-entraîné peuvent être appropriés dans un intervalle de temps donné et deviennent ainsi inappropriés dans un autre intervalle de temps futur sur un même ensemble de données de test ? En effet, ceci s'explique par un changement brusque des données de test après le temps  $t_\alpha$  correspond ainsi à un changement de régime. Donc prédire le temps d'échec d'un modèle pré-entraîné sur un ensemble de données de test équivaut à prédire un changement de régime dans ces données. Le modèle de survie ici agit en prédisant les probabilités de survie du modèle pré-entraîné donc des probabilités de changements de régimes. Dans la section suivante, nous allons discuter des travaux connexes de notre nouveau modèle de survie.

## 2.5 Travaux connexes et contributions

La prédiction d'un changement de régime via la prédiction de la durée de survie d'un modèle pré-entraîné est une méthode nouvelle. À notre connaissance, il n'existe pas encore un modèle similaire à notre modèle de survie. L'article [34] est le seul travail le plus proche de ce mémoire où les auteurs utilisent des modèles de séries temporelles pour la prédiction des pannes et la détection des équipements des réseaux ferroviaires. Notre modèle de survie est différent de ce dernier. En effet, notre modèle prédit la probabilité qu'un modèle déjà entraîné fait des prédictions mieux qu'un modèle naïf, bien défini. Nous allons discuter ici certains travaux connexes avec le modèle de survie même si cette connexion est légère. Comme nous l'avons discuté dans l'introduction, les modèles d'ensemble ont démontré une réussite dans les prédictions. Par exemple, récemment, tous les meilleurs modèles dans la compétition *M4* [77] de séries chronologiques sont des modèles d'ensemble qui appliquent une certaine combinaison entre les différents modèles utilisés [84, 74, 99]. La raison de leur succès est simplement l'utilisation en même temps de plusieurs modèles de prédiction. Ici, chacun de ces modèles fait l'hypothèse sur la distribution de probabilité de l'ensemble d'entraînement [102, 53]. Ainsi, la distribution de probabilité des trajectoires futures

## 2.5. TRAVAUX CONNEXES ET CONTRIBUTIONS

produites par un modèle de prédiction dépend de sa paramétrisation et de l'ensemble des dépendances entre les variables aléatoires. En évaluant leurs performances a posteriori, il est plus facile d'évaluer quelle paramétrage de modèle fonctionne le mieux et d'appliquer une agrégation prévisionnelle du modèle individuel. Cependant, dans la majorité des données chronologiques, cette distribution de probabilité est inconnue ou même intraitable. En pratique, la sélection du modèle à inclure dans le modèle d'ensemble est faite par des méthodes automatiques comme l'optimisation bayésienne [78]. Nous soutenons qu'une telle sélection pourrait être effectuée plus efficacement si nous pouvions mesurer et prédire leur probabilité d'échec à prédire avec précision dans les différents scénarios.

La prédiction de l'échec d'un modèle joue un rôle fondamental en médecine ou dans toutes les situations où une décision basée sur la prédiction d'un modèle doit être prise. Cependant, les applications de ce modèle se font sentir plus dans le monde de la finance et de l'économie où la prédiction des changements de régimes permet de prendre des mesures adéquates dans le but de minimiser les pertes dues. Les contributions de ce mémoire sont les suivantes :

- le modèle de survie est le premier modèle qui permet de prédire les probabilités de survie en fonction du temps des modèles pré-entraînés.
- le modèle de survie est le premier modèle pouvant prédire des changements de régimes en se basant sur la prédiction de l'échec d'un modèle pré-entraînés.
- le modèle de survie peut servir à la détermination des poids dans le modèle FFORMA [74].
- le modèle de survie, comme nous l'avons déjà discuté, peut aider à la sélection des modèles lors de l'entraînement de modèle d'ensemble.

Dans le chapitre suivant, nous allons expérimenter le modèle de survie en utilisant plusieurs modèles pré-entraînés sur des données chronologiques.

# Chapitre 3

## Résultats et discussion

Nous allons expérimenter notre modèle de survie en utilisant un ensemble de modèles pré-entraînés sur des données de séries temporelles, comme décrit dans la sous-section 2.2.1. Cependant, avant de montrer les résultats obtenus, nous décrirons les données chronologiques ayant servies à l’entraînement des modèles pré-entraînés et une brève description sera faite pour chacun de ces modèles.

### 3.1 Données et modèles de séries temporelles

Pour évaluer notre modèle, nous l’avons testé pour prédire l’échec de prédiction des modèles financiers de séries chronologiques. Nous avons construit un ensemble de données de 91663 séries chronologiques univariées à partir des prix de clôture quotidien ajustés de 320 titres observés entre 1990 et 2020 en utilisant un schéma de fenêtre glissante sur une base mensuelle. Ces titres comprennent divers fonds communs de placement américains, des fonds négociés en bourse et des fonds négociés sur les marchés financiers américains, chacun couvrant différents types de classes d’actifs, notamment les actions, les obligations, les matières premières, les devises, les investissements de sociétés et divers autres titres financiers. L’ensemble de données est tiré à partir de trois sources de données : (i) Fast-Track (FT), (ii) La réserve fédérale de Saint-Louis (FRED) et (iii) Yahoo Finance (YF). Chaque série chronologique

### 3.1. DONNÉES ET MODÈLES DE SÉRIES TEMPORELLES

contient au moins 756 observations ou jours pour la formation du modèle de série chronologique individuel et des prévisions ont été effectuées à l'horizon de 21 jours au-delà des 756 jours pour simuler les performances de chaque modèle de série temporelle en prédisant les trajectoires de prix pour le mois à venir. Quelques exemples de ces séries sont montrés dans Fig. 3.1.

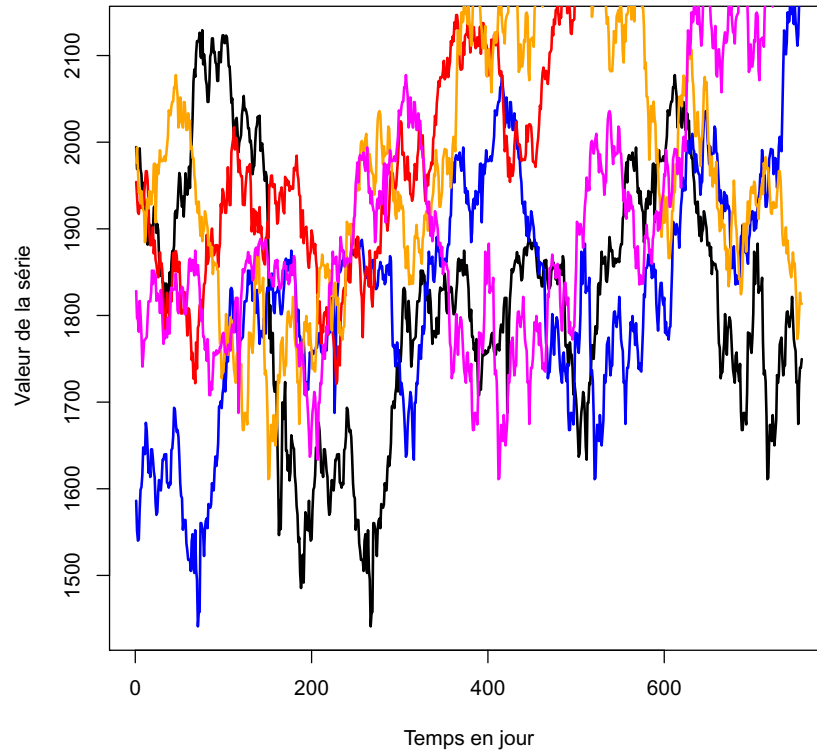


Figure 3.1 – Un ensemble de séries chronologiques tirées des  $N_{en}^u = 91663$  séries d'entraînement.

#### 3.1.1 Les modèles de séries chronologiques

Dans cette sous-section nous allons brièvement décrire les modèles de séries temporelles utilisés pour expérimenter le modèle de survie. Ces modèles seront les mo-

### 3.1. DONNÉES ET MODÈLES DE SÉRIES TEMPORELLES

dèles pré-entraînés. Ce sont la moyenne mobile intégrée auto-régressive (ARIMA) [17, 43, 7], le modèle de marche aléatoire [116, 66, 17], le modèle Prophet [108], le modèle hétéroscédasticité conditionnelle autorégressive généralisée (GARCH) [8], le modèle TBATS (trigonométrique saisonnier, transformation Box-Cox, résidus ARMA, tendance et saisonnalité) [25] et le modèle de réseau de neurones pour les séries chronologiques (NNTAR) [1]. Le modèle **ARIMA**, basé sur le principe de Box-Jenkins [66], est un modèle intuitivement raisonnable pour de nombreuses séries chronologiques qui se produisent dans la pratique. Il est mieux connu sous sa forme plus générale  $\text{ARIMA}(p, d, q)$  où  $p$  est le nombre de termes auto-régressifs,  $d$  le nombre de différences non saisonnières nécessaires pour la stationnarité et  $q$  le nombre de termes à moyenne mobile. Pour une série  $y_t$ , sa formulation mathématique est la suivante :

$$\phi(L)(1 - L)^d y_t = \theta(L)\epsilon_t. \quad (3.1)$$

Dans l'Éq. (3.1),  $\epsilon_t$  est une variable correspondant à l'erreur et normalement distribuée,  $\epsilon_t \sim N(0, \sigma^2)$ .  $L$  est l'opérateur à retardement c'est-à-dire,  $Ly_t = y_{t-1}$  et  $\phi(L)$  et  $\theta(L)$  sont des fonctions polynomiales de  $L$ . Ils s'expriment sous la forme :

$$\phi(L) = 1 - \sum_{i=1}^p \phi_i L^i \quad \text{et} \quad \theta(L) = 1 - \sum_{j=1}^q \theta_j L^j, \quad (3.2)$$

où  $\phi_i$  et  $\theta_j$  sont les paramètres du modèle. Selon les valeurs des constantes  $p, q$  et  $d$ , cette formulation de modèle général peut donner naissance à plusieurs modèles particuliers :

- $\text{ARIMA}(p, 0, q)$  est le modèle de la moyenne mobile d'auto-régressive ( $\text{ARMA}(p, q)$ ) [17]
- $\text{ARIMA}(p, 0, 0)$  est le modèle auto-régressif  $\text{AR}(p)$  [66, 7]
- $\text{ARIMA}(0, 0, q)$  est le modèle de la moyenne mobile  $\text{MA}(p)$  [7, 43]
- le **modèle à marche aléatoire** que nous avons utilisé ici correspond au modèle  $\text{ARIMA}(0, 1, 0)$ .

**Prophet** est un modèle de séries chronologiques permettant de prendre en compte les caractéristiques communes aux séries financières telles que les fortes saisonnalités multiples, les changements de tendance, les valeurs aberrantes et les effets de vacances.



### 3.1. DONNÉES ET MODÈLES DE SÉRIES TEMPORELLES

Mathématiquement, une série  $y_t$  est décomposée en ces caractéristiques sous la forme :

$$y_t = g(t) + s(t) + h(t) + \epsilon_t. \quad (3.3)$$

Dans l'Éq. (3.3),  $g(t)$  représente les changements non périodiques de la série,  $s(t)$  les changements périodiques, par exemple la saisonnalité annuelle, et  $h(t)$  les effets de vacances.  $\epsilon_t$  est le terme d'erreur et représente tout changement idiosyncratique qui n'est pas pris en compte par le modèle.

**GARCH** est un modèle de moyenne mobile auto-régressive où une variance d'erreur est également supposée en plus des termes d'erreur. C'est un modèle qui peut être utilisé pour différents types de données de séries chronologiques financières. Le modèle GARCH( $p, q$ ), avec  $p \geq 0$  et  $q > 0$ , s'exprime comme :

$$y_t | \psi_{t-1} \sim N(0, h_t) \quad \text{avec} \quad h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i}. \quad (3.4)$$

où  $\alpha_i$  et  $\beta_i$  sont les paramètres du modèle et  $\psi_{t-1}$  contient les valeurs historiques au temps  $t-1$ . Si  $p = 0$ , le modèle GARCH( $p = 0, q$ ) se réduit au modèle hétéroscédasticité conditionnelle auto-régressive ARCH( $q$ ). Dans le processus ARCH( $q$ ), la variance conditionnelle est spécifiée comme une fonction linéaire des variances d'échantillons historiques uniquement, tandis que le processus GARCH( $p, q$ ) tient également en compte les variances conditionnelles décalées.

Le modèle **BATS** est le modèle dite Holt-Winters [105] où on inclut la transformation de Box – Cox, les erreurs du modèle de ARIMA et les patrons saisonniers de

### 3.1. DONNÉES ET MODÈLES DE SÉRIES TEMPORELLES

période  $T$ . Mathématiquement on peut l'écrire sous la forme :

$$\begin{aligned}
y_t^{(w)} &= \begin{cases} \frac{y_t^{(w)} - 1}{w} & \text{si } w \neq 0 \\ \log y_t & \text{si } w = 0, \end{cases} \\
y_t^{(w)} &= l_{t-1} + \phi b_{t-1} + \sum_{i=1}^T s_{t-m_i}^{(i)} + d_t, \\
l_t &= l_{t-1} + \phi b_{t-1} + \alpha d_t, \\
b_t &= (1 - \phi)b + \phi b_{t-1} + \beta d_t, \\
s_t^{(i)} &= s_{t-m_i}^{(i)} + d_t, \\
d_t &= \sum_{i=1}^p \phi_i d_{t-i} + \sum_{i=1}^q \theta_i \epsilon_{t-i} + \epsilon_t,
\end{aligned}$$

où les composantes  $l_t$  et  $b_t$  représentent respectivement les composantes de niveau et de tendance de la série, et  $s_t^{(i)}$  la  $i$ ème composante saisonnière au temps  $t$ . Les coefficients  $\alpha$ ,  $\beta$  sont les paramètres de lissage,  $b$  est la tendance à long terme et  $b_t$  est la tendance à court terme de la période  $t$ .  $d_t$  désigne un processus ARMA( $p, q$ ) et  $\epsilon_t \sim N(0, \sigma^2)$ .  $\phi$  est un paramètre d'amortissement. Si les composantes saisonnières  $s_t^{(i)}$  sont remplacées par une formulation saisonnière trigonométrique [35], et l'équation de mesure par :

$$y_t^{(w)} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^T s_{t-1}^{(i)} + d_t, \quad (3.5)$$

on obtient le modèle dit TBATS.

En fin, le modèle **NNTAR** utilisé ici est une application des réseaux de neurones, décrits dans la section 1.3, aux séries temporelles.

Tous les modèles précités ont été entraînés avec l'ensemble d'entraînement  $N_{tr}^u$ . Après l'entraînement de ces modèles, nous avons adopté un processus d'extraction des caractéristiques comme décrit dans section 2.2.3 et un processus de génération du processus de survie (voir section 2.2.2). On obtient ainsi l'ensemble des données d'entraînement pour le modèle de survie dont les processus de survie déterminés correspondent aux cibles.

### 3.2. RÉSULTATS

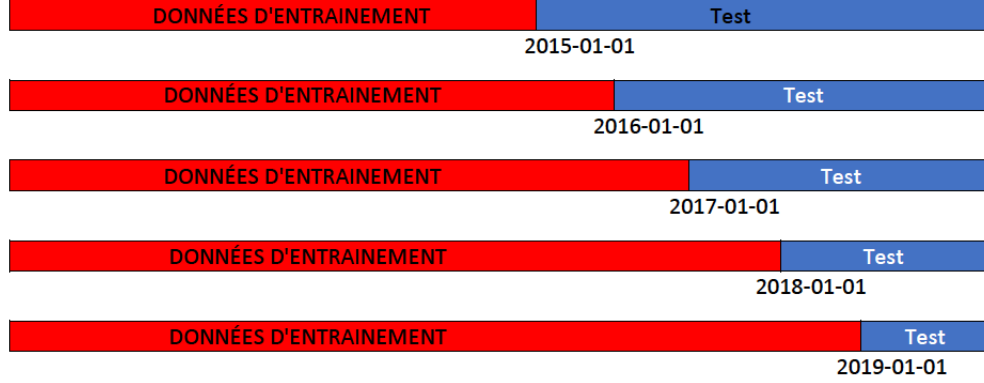


Figure 3.2 – La segmentation des données en ensemble d’entraînement et de test.

## 3.2 Résultats

Dans la figure 3.2, on montre les différentes segmentations que nous avons adoptées. L’ensemble d’entraînement  $N_{tr}$  augmente annuellement de 2015 à 2019, contrairement, l’ensemble de test  $N_{te}$  qui diminue pour chaque année. Cette procédure permettra de comprendre comment l’erreur faite par le modèle de survie sur les données de test évolue au fur du temps et de savoir en plus si cette erreur est stable. En fait, on peut voir cette façon de faire comme une validation croisée mais dont on doit respecter la chronologie des données. Suivant cette segmentation, nous avons trouvé les paramètres du modèle de survie, décrits à la section 2.3.

La figure 3.4 montre les probabilités prédites par le modèle de survie sur l’ensemble de test. Elle représente les résultats de prédictions du modèle de survie sur les différents modèles pré-entraînés. Dans cette figure, nous avons montré le processus de survie à gauche (où les cibles) et les probabilités prédites correspondantes  $\mathcal{P}_s(t_\alpha|\mathbf{X}, \mathbf{W}, \mathbf{b})$ ,  $s = 1, 2, \dots, N_{te}^u$  à droite pour chacune des séries dans l’ensemble de test. Les probabilités donnent les pourcentages pour qu’un modèle pré-entraîné prédise mieux que le modèle de référence. Dans la figure 3.4, il est important de noter que nous avons choisi aléatoirement dix séries parmi les  $N_{te}$  utilisées dans l’ensemble de test. Les valeurs du processus de survie appartiennent à l’ensemble  $\{1, 0\}$  contrairement aux probabilités prédites qui sont des distributions continues. Comme la perfor-

### 3.2. RÉSULTATS

mance de prédiction d'un modèle pré-entraîné dépend principalement des propriétés des séries temporelles, par exemple, la saisonnalité etc., nous obtenons des processus de survie différents spécifiques pour chacune des séries et modèles pré-entraînés. Néanmoins, pour une série et un modèle pré-entraîné donnés, on peut remarquer facilement que les courbes de probabilités  $\mathcal{P}_s(t_\alpha|\mathbf{X}, \mathbf{W}, \mathbf{b})$ ,  $s = 1, 2, \dots, N_{te}^u$  suivent en bonne partie le comportement des courbes représentant le processus de survie. Cependant, il existe des séries pour lesquelles la probabilité de survie prédite pour le modèle pré-entraîné diminue plus rapidement comparée au processus de survie correspondant. Ceci implique que le modèle de survie prédit l'échec du modèle pré-entraîné plus tôt que prévu. Notez que les probabilités de sortie nous informent sur les performances du modèle pré-entraîné par rapport au modèle de référence, c'est-à-dire, la probabilité que le modèle pré-entraîné soit plus précis à prédire la réalité que le modèle de référence pour tout instant  $t$  donné. Évidemment, cette probabilité est égale à 1 à  $t = 0$  et diminue lorsque le nombre de jours augmente, comme prévu théoriquement.

#### 3.2.1 Évaluation du modèle de survie

Pour évaluer la performance du modèle de survie sur l'ensemble de test, nous avons mesuré des métriques comme le CBS, le MAE (adaptée à l'analyse de survie) et le C-index données respectivement par les équations (1.30), (1.32) et (1.28), décrites dans section 1.1.5. Le tableau 3.1 montre les valeurs du CBS pour les séquences des données décrites par la figure 3.2 et pour chacun des modèles pré-entraînés. En utilisant la métrique CBS pour évaluer notre modèle, on dira que le modèle de survie est précis si la valeur du CBS est petite. Généralement, une valeur du CBS inférieure à 0.5 est une bonne indication quant à la bonne performance du modèle de prédiction. Selon l'interprétation du CBS, on observe bien que le modèle de survie a bien performé en prédisant la durée de survie des différentes modèles pré-entraînés : la valeur moyenne du CBS observée, pour chaque modèle pré-entraîné est inférieure à 0.4. Cependant, le modèle à marche aléatoire RW correspondant au modèle ARIMA(0, 1, 0) a la valeur du CBS la plus petite. Ainsi, le modèle de survie serait plus performant à prédire sa durée de survie, c'est-à-dire, à prédire pour combien de temps  $t$  avec quelle probabilité le modèle RW sera plus performant que le modèle de référence. Toujours

### 3.2. RÉSULTATS

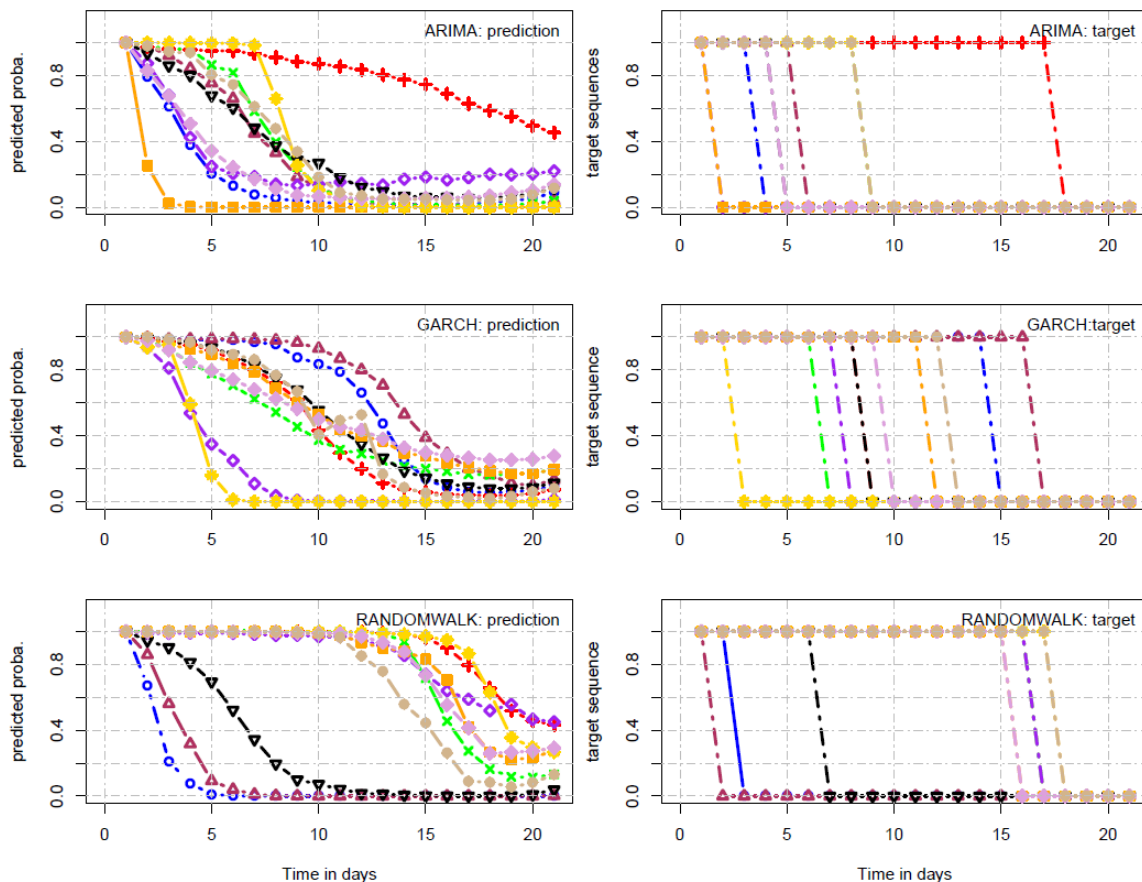


Figure 3.3 – Probabilité de survie à gauche et le processus de survie à droite pour dix séries chronologiques sélectionnées aléatoirement. Les probabilités donnent le pourcentage que les modèles pré-entraînés **ARIMA**, **GARCH** et le **modèle marche aléatoire (RW)** performant mieux que le modèle de référence, c'est-à-dire le modèle naïf, sur l'ensemble des données de test. Comme nous l'avons déjà discuter précédemment, le processus de survie devient 1 si le modèle pré-entraîné est plus performant que le modèle naïf sur l'ensemble de test  $N_{te}^u$  et 0 dans le cas contraire. Le processus de survie prend donc seulement deux valeurs : 1 ou 0. Les probabilités de survie sont quant à elles des distributions continues. Les triangles, carrés etc. n'ont aucune signification particulière. Chaque courbe correspond à la probabilité prédite pour une série temporelle donnée.

### 3.2. RÉSULTATS

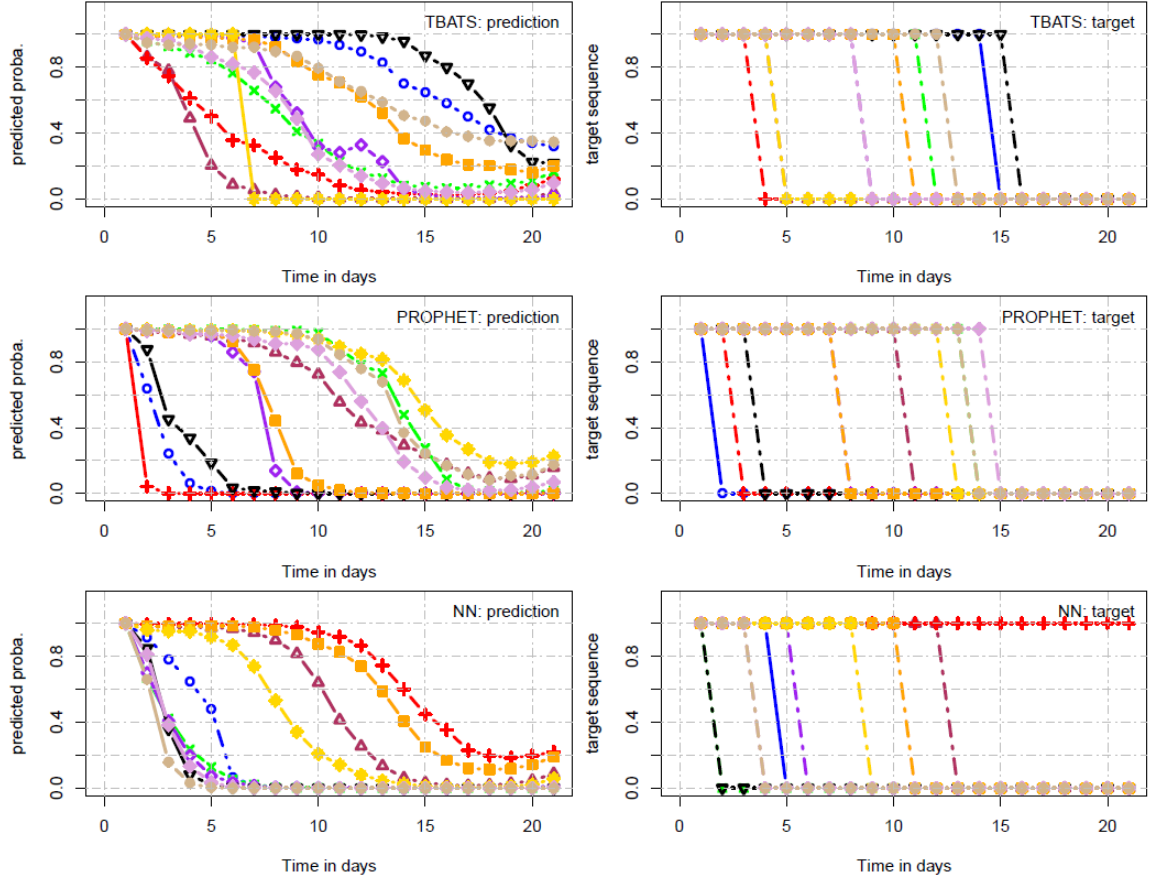


Figure 3.4 – Probabilité de survie à gauche et le processus de survie à droite pour dix séries chronologiques sélectionnées aléatoirement. Les probabilités donnent le pourcentage que les modèles pré-entraînés **TBATS**, **PROPHET** et **NN** performent mieux que le modèle de référence, c'est-à-dire le modèle naïf, sur l'ensemble des données de test. Comme nous l'avons déjà discuter précédemment, le processus de survie devient 1 si le modèle pré-entraîné est plus performant que le modèle naïf sur l'ensemble de test  $N_{te}^u$  et 0 dans le cas contraire. Le processus de survie prend donc seulement deux valeurs : 1 ou 0. Les probabilités de survie sont quant à elles des distributions continues. Les triangles, carrés etc. n'ont aucune signification particulière. Chaque courbe correspond à la probabilité prédite pour une série temporelle donnée.

### 3.2. RÉSULTATS

Tableau 3.1 – Les valeurs de la métrique CBS trouvées en évaluant le modèle de survie pour chacun des modèles pré-entraîné : ARIMA, GARCH, PROPHET, TBATS, RW et NNETAR. Le modèle de survie a été évalué selon la segmentation des données montrée dans la figure 3.2. La dernière valeur de chaque sous-tableau correspond à la moyenne des cinq différentes segmentations.

	ARIMA	GARCH	PROPHET
CBS (2015)	0.289	0.296	0.322
CBS (2016)	0.274	0.269	0.346
CBS (2017)	0.260	0.261	0.330
CBS (2018)	0.266	0.262	0.313
CBS (2019)	0.230	0.221	0.312
CBS (moyenne)	0.265	0.261	0.324
	TBATS	RW	NNTAR
CBS (2015)	0.283	0.256	0.361
CBS (2016)	0.280	0.237	0.360
CBS (2017)	0.284	0.240	0.354
CBS (2018)	0.273	0.265	0.335
CBS (2019)	0.278	0.214	0.363
CBS (moyenne)	0.279	<b>0.242</b>	0.354

selon cette métrique, on note aussi que le modèle de survie performe moins lorsqu'on essaie de prédire la durée de survie du modèle pré-entraîné NNTAR. En effet, la valeur du CBS la plus élevée est obtenue avec ce modèle NNTAR.

Le tableau 3.2 présente les résultats obtenus en mesurant la métrique MAE . Comme dans le cas du CBS, une valeur faible du MAE correspond à une bonne performance du modèle de survie à prédire la durée de survie d'un modèle pré-entraîné. Nous constatons que toutes les valeurs du MAE sont petites pour tous les différents modèles pré-entraînés ce qui informe sur la bonne performance du modèle de survie.

Enfin, le tableau 3.3 donne les valeurs de la métrique C-index. Pour le C-index, des valeurs élevées signifient que le modèle prédit des probabilités de survie plus élevées pour des durées de survie observées plus élevées. Ainsi, un modèle avec des valeurs élevées du C-index indique une bonne performance de notre modèle de survie à prédire les probabilités de survie des modèles pré-entraînés. On constate que toutes les valeurs du C-index sont autour de 0.6. Ainsi, la métrique C-index démontre à nouveau que le modèle de survie a une bonne performance à prédire la durée de survie des modèles pré-entraînés.

### 3.2. RÉSULTATS

Tableau 3.2 – Les valeurs du MAE trouvées en évaluant le modèle de survie pour chacun des modèles pré-entraînés : ARIMA, GARCH, PROPHET, TBATS, RW et NNTAR. Le modèle de survie a été évalué selon la segmentation des données montrée dans la figure 3.2. La dernière valeur de chaque sous-tableau correspond à la moyenne des cinq différentes segmentations.

	ARIMA	GARCH	PROPHET
MAE (2015)	0.195	0.199	0.201
MAE (2016)	0.189	0.186	0.211
MAE (2017)	0.188	0.185	0.209
MAE (2018)	0.189	0.187	0.200
MAE (2019)	0.173	0.170	0.202
MAE (moyenne)	0.186	0.185	0.204
	TBATS	RW	NNTAR
MAE (2015)	0.194	0.179	0.225
MAE (2016)	0.195	0.171	0.226
MAE (2017)	0.198	0.173	0.223
MAE (2018)	0.191	0.187	0.219
MAE (2019)	0.195	0.15	0.229
MAE (moyenne)	0.194	<b>0.172</b>	0.224

Tableau 3.3 – Les valeurs du C-index trouvées en évaluant le modèle de survie pour chacun des modèles pré-entraînés : ARIMA, GARCH, PROPHET, TBATS, RW et NNTAR. Le modèle de survie a été évalué selon la segmentation des données montrée dans Fig. 3.2. La dernière valeur de chaque sous-tableau correspond à la moyenne des cinq différentes segmentations.

	ARIMA	GARCH	PROPHET
C-index (2015)	0.619	0.612	0.647
C-index (2016)	0.629	0.615	0.635
C-index (2017)	0.594	0.590	0.610
C-index (2018)	0.606	0.602	0.648
C-index (2019)	0.582	0.570	0.637
C-index (moyenne)	0.606	0.598	0.635
	TBATS	RW	NNTAR
C-index (2015)	0.626	0.668	0.621
C-index (2016)	0.618	0.652	0.608
C-index (2017)	0.605	0.644	0.618
C-index (2018)	0.620	0.643	0.614
C-index (2019)	0.607	0.684	0.616
C-index (moyenne)	0.615	<b>0.658</b>	0.615



### 3.2. RÉSULTATS

En résumé, nous remarquons que les trois métriques utilisées démontrent une bonne capacité du modèle de survie à prédire la durée de survie des modèles pré-entraînés.

# Conclusion

Dans ce mémoire nous avons élaboré un nouveau modèle de survie pour les séries chronologiques capable de prédire le temps où un modèle prédictif pré-entraîné et performant au début devient moins performant qu'un modèle naïf. Le nouveau modèle de survie fournit précisément les probabilités, en fonction du temps, que le modèle prédictif pré-entraîné soit plus performant qu'un modèle naïf et donc, quand nous ne devons plus faire confiance aux prédictions du modèle prédictif pré-entraîné. Notre modèle de survie peut être aussi utilisé dans la détection des changements de régimes dans les séries chronologiques et financières en particulier. En effet, nous pouvons quelque fois associer les échecs des modèles prédictifs déjà bien entraînés à des changements importants dans les séries chronologiques. Cependant, nous devons préciser que les échecs des modèles prédictifs ne sont pas seulement dûs à des changements de régimes, ils peuvent être causés par d'autres propriétés des séries temporelles comme la saisonnalité.

Notre modèle de survie comporte quatre phases. On utilise tout d'abord les techniques d'entraînement et de test bien connues pour entraîner et tester la généralisation des modèles de séries chronologiques que nous désirons vérifier la durée de leur performance. Ensuite, on entame la génération d'un processus de survie en testant ces modèles pré-entraînés sur un ensemble de test et en comparant leur performance avec celle d'un modèle naïf. Pour une série chronologique donnée, le processus de survie sera de 1 au temps  $t$  si le modèle pré-entraîné est plus performant que le modèle naïf 0 sinon, et constitue les cibles pour l'entraînement du nouveau modèle de survie. Dans la troisième phase du modèle de survie, un ensemble de caractéristiques est extrait à partir de l'ensemble des séries permettant d'entraîner les modèles de séries temporelles. Ces caractéristiques seront les données d'entraînement du modèle de survie en

## CONCLUSION

plus du processus de survie. La dernière phase du modèle de survie est son entraînement, avec un réseau de neurones profonds, en minimisant une entropie croisée avec la méthode d'Adam. La généralisation du modèle de survie sur l'ensemble de test permet de donner en sortie une distribution de probabilité en tout temps  $t$  indiquant avec quelle probabilité, le modèle pré-entraîné est plus performant qu'un modèle naïf.

Nous avons obtenu la distribution de probabilités de survie de chaque modèle pré-entraîné en utilisant des données financières. Une comparaison avec les vraies cibles nous permet de conclure que les probabilités de survie suivent en bonne partie les processus de survie indiquant ainsi la performance du nouveau modèle de survie à prédire la survie des modèles pré-entraînés. Nous avons évalué notre modèle de survie en utilisant des mesures métriques bien adaptées à l'analyse de survie telles que le CBS, le MAE (adaptée à l'analyse de survie) et le C-index. Toutes ces métriques indiquent une bonne performance du modèle de survie à prédire la survie des modèles pré-entraînés. Les contributions de ce mémoire sont donc les suivantes :

(i) Notre modèle de survie est le premier qui permet de prédire les probabilités de survie en fonction du temps de modèles de séries chronologiques pré-entraînés.

(ii) Le modèle de survie est aussi le premier modèle pouvant prédire un changement de régime en se basant sur la prédiction de l'échec d'un modèle de séries temporelles pré-entraînés. La prédiction de l'échec d'un modèle joue un rôle fondamental en médecine ou dans toutes les situations où une décision basée sur la prédiction d'un modèle doit être prise. Cependant, les applications de ce modèle se font sentir plus dans le monde de la finance et de l'économie où la prédiction de changement de régime permet de prendre des mesures adéquates dans le but de minimiser les pertes dues à ces changements de régimes.

# Bibliographie

- [1] R. Adhikari et R. K. Agrawal, « An introductory study on time series modeling and forecasting, » *arXiv preprint arXiv :1302.6613*, 2013.
- [2] O. Aalen, « Nonparametric inference for a family of counting processes, » *The Annals of Statistics*, pp. 701–726, 1978.
- [3] S. Ameri, M. J. Fard, R. B. Chinnam, et C. K. Reddy, « Survival analysis based framework for early prediction of student dropouts, » dans *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 2016, pp. 903–912.
- [4] P. K. Andersen et R. D. Gill, « Cox’s regression model for counting processes : a large sample study, » *The annals of statistics*, pp. 1100–1120, 1982.
- [5] K. Ajay, R. Sushil, et A. Tiwari, « Cancer Survival Analysis Using Machine Learning, » *Available at SSRN 3354469*, 2019.
- [6] I. Bou-Hamad, D. Larocque, H. Ben-Ameur *et al.*, « A review of survival trees, » *Statistics Surveys*, vol. 5, pp. 44–71, 2011.
- [7] G. E. Box, G. M. Jenkins, et G. Reinsel, « Time series analysis : forecasting and control Holden-day San Francisco, » *BoxTime Series Analysis : Forecasting and Control Holden Day1970*, 1970.
- [8] T. Bollerslev, « Generalized autoregressive conditional heteroskedasticity, » *Journal of econometrics*, vol. 31, no. 3, pp. 307–327, 1986.
- [9] G. W. Brier, « Verification of forecasts expressed in terms of probability, » *Monthly weather review*, vol. 78, no. 1, pp. 1–3, 1950.

## BIBLIOGRAPHIE

- [10] N. Baba et Y. Sakurai, « Predicting regime switches in the VIX index with macroeconomic variables, » *Applied Economics Letters*, vol. 18, no. 15, pp. 1415–1419, 2011.
- [11] R. Bellazzi et B. Zupan, « Predictive data mining in clinical medicine : current issues and guidelines, » *International journal of medical informatics*, vol. 77, no. 2, pp. 81–97, 2008.
- [12] A. Ciampi, R. Bush, M. Gospodarowicz, et J. Till, « An approach to classifying prognostic factors related to survival experience for non-Hodgkin’s lymphoma patients : Based on a series of 982 patients : 1967–1975, » *Cancer*, vol. 47, no. 3, pp. 621–627, 1981.
- [13] A. Ciampi, C.-H. Chang, S. Hogg, et S. McKinney, « Recursive partition : A versatile method for exploratory-data analysis in biostatistics, » dans *Biostatistics*. Springer, 1987, pp. 23–50.
- [14] C. Cappelli, P. D’Urso, et F. Di Iorio, « Regime change analysis of interval-valued time series with an application to PM10, » *Chemometrics and Intelligent Laboratory Systems*, vol. 146, pp. 337–346, 2015.
- [15] S. J. Cutler et F. Ederer, « Maximum utilization of the life table method in analyzing survival, » *Journal of chronic diseases*, vol. 8, no. 6, pp. 699–712, 1958.
- [16] C. L. Chiang, « The life table and its applications, » 1984.
- [17] J. H. Cochrane, « Time series for macroeconomics and finance, » *Manuscript, University of Chicago*, pp. 1–136, 2005.
- [18] D. Collett, *Modelling survival data in medical research*. CRC press, 2015.
- [19] D. R. Cox, « Regression models and life-tables, » *Journal of the Royal Statistical Society : Series B (Methodological)*, vol. 34, no. 2, pp. 187–202, 1972.
- [20] D. R. Cox, « Partial likelihood, » *Biometrika*, vol. 62, no. 2, pp. 269–276, 1975.

## BIBLIOGRAPHIE

- [21] C.-L. Chi, W. N. Street, et W. H. Wolberg, « Application of artificial neural network-based survival analysis on two breast cancer datasets, » dans *AMIA Annual Symposium Proceedings*, vol. 2007. American Medical Informatics Association, 2007, p. 130.
- [22] A. Ciampi, J. Thiffault, J.-P. Nakache, et B. Asselain, « Stratification by stepwise regression, correspondence analysis and recursive partition : a comparison of three methods of analysis for survival data with co-variates, » *Computational statistics & data analysis*, vol. 4, no. 3, pp. 185–204, 1986.
- [23] R. B. Davis et J. R. Anderson, « Exponential survival trees, » *Statistics in Medicine*, vol. 8, no. 8, pp. 947–961.
- [24] J. G. De Gooijer et R. J. Hyndman, « 25 years of time series forecasting, » *International journal of forecasting*, vol. 22, no. 3, pp. 443–473, 2006.
- [25] A. M. De Livera, R. J. Hyndman, et R. D. Snyder, « Forecasting time series with complex seasonal patterns using exponential smoothing, » *Journal of the American statistical association*, vol. 106, no. 496, pp. 1513–1527, 2011.
- [26] M. L. De Prado, *Advances in financial machine learning*. John Wiley & Sons, 2018.
- [27] A. Davis, Roger B et J. R., « Exponential survival trees, » *Statistics in Medicine*, vol. 8, no. 8, pp. 947–961, 1989.
- [28] S. Y. D. Dubey, « Normal and Weibull distributions, » *Naval Research Logistics Quarterly*, vol. 14, no. 1, pp. 69–79, 1967.
- [29] O. J. Dunn, « Basic Statistics. A Primer for the Biomedical Sciences, » *Academic Medicine*, vol. 40, no. 1, p. 70, 1965.
- [30] B. D. Fulcher, « Feature-based time-series analysis, » *arXiv preprint arXiv :1709.08055*, 2017.
- [31] I. Goodfellow, Y. Bengio, et A. Courville, « Machine learning basics, » *Deep learning*, vol. 1, pp. 98–164, 2016.

## BIBLIOGRAPHIE

- [32] E. A. Gehan, « Estimating survival functions from the life table, » *Journal of Clinical Epidemiology*, vol. 21, no. 9, pp. 629–644, 1969.
- [33] L. Gordon et R. A. Olshen, « Tree-structured survival analysis. » *Cancer treatment reports*, vol. 69, no. 10, pp. 1065–1069, 1985.
- [34] F. P. García, D. J. Pedregal, et C. Roberts, « Time series methods applied to failure prediction and detection, » *Reliability Engineering and System Safety*, vol. 95, no. 6, pp. 698–703, 2010.
- [35] A. C. Harvey, « Forecasting, structural time series models and the Kalman filter, » 1990.
- [36] T. Hesterberg, N. H. Choi, L. Meier, C. Fraley *et al.*, « Least angle and  $\ell_1$  penalized regression : A review, » *Statistics Surveys*, vol. 2, pp. 61–93, 2008.
- [37] F. E. Harrell, R. M. Califf, D. B. Pryor, K. L. Lee, et R. A. Rosati, « Evaluating the yield of medical tests, » *Jama*, vol. 247, no. 18, pp. 2543–2546, 1982.
- [38] M. A. Hernán, « The hazards of hazard ratios, » *Epidemiology, Cambridge, Mass*, vol. 21, no. 1, p. 13, 2010.
- [39] F. E. Harrell Jr, K. L. Lee, R. M. Califf, D. B. Pryor, et R. A. Rosati, « Regression modelling strategies for improved prognostic prediction, » *Statistics in medicine*, vol. 3, no. 2, pp. 143–152, 1984.
- [40] A. E. Hoerl et R. W. Kennard, « Ridge Regression : Biased Estimation for Nonorthogonal Problems, » *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.
- [41] R. Hyndman, Y. Kang, P. Montero-Manso, T. Talagala, E. Wang, Y. Yang, et M. O’Hara-Wild, « tsfeatures : Time Series Feature Extraction, » *R package version*, vol. 1, no. 0, 2019.
- [42] T. Hothorn, B. Lausen, A. Benner, et M. Radespiel-Tröger, « Bagging survival trees, » *Statistics in medicine*, vol. 23, no. 1, pp. 77–91, 2004.
- [43] K. W. Hipel et A. I. McLeod, « Time series modelling of water resources and environmental systems, » 1994.

## BIBLIOGRAPHIE

- [44] S. Har-Peled, D. Roth, et D. Zimak, « Constraint classification for multiclass classification and ranking, » pp. 809–816, 2003.
- [45] B. M. Henrique, V. A. Sobreiro, et H. Kimura, « Literature review : Machine learning techniques applied to financial market prediction, » *Expert Systems with Applications*, vol. 124, pp. 226–251, 2019.
- [46] V. Im et M. Briex, « Médecine prédictive, deep learning, algorithmes et accouchement, » *Spirale*, no. 1, pp. 204–209, 2020.
- [47] K. J. Ishak, N. Kreif, A. Benedict, et N. Muszbek, « Overview of parametric survival analysis for health-economic applications, » *Pharmacoeconomics*, vol. 31, no. 8, pp. 663–675, 2013.
- [48] P. C. Jackson, « Introduction to artificial intelligence, » 2019.
- [49] S. P. Jenkins, « Survival analysis, » *Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK*, vol. 42, pp. 54–56, 2005.
- [50] S. P. Jenkins, « Survival analysis, » *Unpublished manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK*, vol. 42, pp. 54–56, 2005.
- [51] S. Johansen, « An extension of Cox’s regression model, » *International Statistical Review*, pp. 165–174, 1983.
- [52] D. P. Kingma et J. Ba, « Adam : A method for stochastic optimization, » *arXiv preprint arXiv :1412.6980*, 2014.
- [53] D. Koller et N. Friedman, « Probabilistic graphical models : principles and techniques, » 2009.
- [54] D. Kleinbaum et M. Klein, « Survival Analysis : A Self-Learning Text, » *Biometrics*, vol. 62, no. 1, pp. 312–312, 2006.
- [55] E. L. Kaplan et P. Meier, « Nonparametric estimation from incomplete observations, » *Journal of the American statistical association*, vol. 53, no. 282, pp. 457–481, 1958.
- [56] J. P. Klein et M. L. Moeschberger, « Survival analysis : techniques for censored and truncated data, » 2006.



## BIBLIOGRAPHIE

- [57] B. Krawczyk, L. L. Minku, J. Gama, J. Stefanowski, et M. Woźniak, « Ensemble learning for data stream analysis : A survey, » *Information Fusion*, vol. 37, pp. 132–156, 2017.
- [58] N. Kourentzes, F. Petropoulos, et J. R. Trapero, « Improving forecasting by estimating time series structural components across multiple frequencies, » *International Journal of Forecasting*, vol. 30, no. 2, pp. 291–302, 2014.
- [59] B. Krollner, B. J. Vanstone, et G. R. Finnie, « Financial time series forecasting with machine learning techniques : a survey. » dans *ESANN*, 2010.
- [60] F. M. Khan et V. B. Zubek, « Support vector regression for censored data (SVRc) : a novel tool for survival analysis, » dans *2008 Eighth IEEE International Conference on Data Mining*. IEEE, 2008, pp. 863–868.
- [61] L. Liao et H.-i. Ahn, « Combining deep learning and survival analysis for asset health management, » *Int J Progn Health Manag*, vol. 7, p. 020, 2016.
- [62] N. R. Latimer, « Survival analysis for economic evaluations alongside clinical trials—extrapolation with patient-level data : inconsistencies, limitations, and a practical guide, » *Medical Decision Making*, vol. 33, no. 6, pp. 743–754, 2013.
- [63] Y. LeCun, Y. Bengio, et G. Hinton, « Deep learning, » *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [64] M. LeBlanc et J. Crowley, « Relative risk trees for censored survival data, » *Biometrics*, pp. 411–425, 1992.
- [65] M. LeBlanc et J. Crowley, « Survival trees by goodness of split, » *Journal of the American Statistical Association*, vol. 88, no. 422, pp. 457–467, 1993.
- [66] J. Lee, « Univariate time series modeling and forecasting (Box-Jenkins method), Econ 413, Lecture 4, » *Department of Economics, University of Illinois*, 2018.

## BIBLIOGRAPHIE

- [67] W.-Y. Lin, Y.-H. Hu, et C.-F. Tsai, « Machine learning in financial crisis prediction : a survey, » *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 421–436, 2011.
- [68] D. Lin, « On the Breslow estimator, » *Lifetime data analysis*, vol. 13, no. 4, pp. 471–480, 2007.
- [69] E. T. Lee et J. Wang, « Statistical methods for survival data analysis, » vol. 476, 2003.
- [70] Y. Li, K. S. Xu, et C. K. Reddy, « Regularized parametric regression for high-dimensional survival analysis, » dans *Proceedings of the 2016 SIAM International Conference on Data Mining*. SIAM, 2016, pp. 765–773.
- [71] S. Makridakis, R. J. Hyndman, et F. Petropoulos, « Forecasting in social settings : The state of the art, » *International Journal of Forecasting*, vol. 36, no. 1, pp. 15–28, 2020.
- [72] U. B. Mogensen, H. Ishwaran, et T. A. Gerds, « Evaluating random forests for survival analysis using prediction error curves, » *Journal of statistical software*, vol. 50, no. 11, p. 1, 2012.
- [73] R. G. Miller Jr, *Survival analysis*. John Wiley & Sons, 2011, vol. 66.
- [74] P. Montero-Manso, G. Athanasopoulos, R. J. Hyndman, et T. S. Talagala, « FFORMA : Feature-based forecast model averaging, » *International Journal of Forecasting*, vol. 36, no. 1, pp. 86–92, 2020.
- [75] R. Mitchell, J. Michalski, et T. Carbonell, « An artificial intelligence approach, » 2013.
- [76] S. Mullainathan et J. Spiess, « Machine learning : an applied econometric approach, » *Journal of Economic Perspectives*, vol. 31, no. 2, pp. 87–106, 2017.
- [77] S. Makridakis, E. Spiliotis, et V. Assimakopoulos, « The M4 Competition : Results, findings, conclusion and way forward, » *International Journal of Forecasting*, vol. 34, no. 4, pp. 802–808, 2018.

## BIBLIOGRAPHIE

- [78] M. Moritani et Y. Takefuji, « Ensemble machine learning based on a voting method using red wine dataset for business education, » *Alcohol*, vol. 8, no. 14.90, pp. 10–42.
- [79] E. Marubini et M. G. Valsecchi, « Analysing survival data from clinical trials and observational studies, » vol. 15, 2004.
- [80] W. Nelson, « Theory and applications of hazard plotting for censored failure data, » *Technometrics*, vol. 14, no. 4, pp. 945–966, 1972.
- [81] W. Nelson, « Theory and applications of hazard plotting for censored failure data, » *Technometrics*, vol. 42, no. 1, pp. 12–25, 2000.
- [82] N. J. Nilsson, « Principles of artificial intelligence, » 2014.
- [83] G. H. Oliveira, R. C. Cavalcante, G. G. Cabral, L. L. Minku, et A. L. Oliveira, « Time series forecasting in the presence of concept drift : A pso-based approach, » dans *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 2017, pp. 239–246.
- [84] B. N. Oreshkin, D. Carpov, N. Chapados, et Y. Bengio, « N-BEATS : Neural basis expansion analysis for interpretable time series forecasting, » *arXiv preprint arXiv :1905.10437*, 2019.
- [85] S. M. Pincus, « Approximate entropy as a measure of system complexity. » *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, 1991.
- [86] A. V. Peterson Jr, « Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions, » *Journal of the American Statistical Association*, vol. 72, no. 360a, pp. 854–858, 1977.
- [87] F. Provost, « Toward economic machine learning and utility-based data mining, » dans *Proceedings of the 1st international workshop on Utility-based data mining*, 2005, pp. 1–1.
- [88] U. Ray, U. Chouhan, et N. Verma, « Comparative study of machine learning approaches for classification and prediction of selective caspase-3 antagonist for Zika virus drugs, » *Neural Computing and Applications*, pp. 1–18, 2020.

## BIBLIOGRAPHIE

- [89] D. E. Rumelhart, G. E. Hinton, et R. J. Williams, « Learning representations by back-propagating errors, » *nature*, vol. 323, no. 6088, pp. 533–536, 1986.
- [90] S. Russell et P. Norvig, « Artificial intelligence : a modern approach, » 2002.
- [91] S. N. Rodionov, « A sequential method for detecting regime shifts in the mean and variance, » *Large-scale disturbances (regime shifts) and recovery in aquatic ecosystems : challenges for management toward sustainability*, pp. 68–72, 2005.
- [92] M. V. Shcherbakov, A. Brebels, N. L. Shcherbakova, A. P. Tyukov, T. A. Janovsky, et V. A. Kamaev, « A survey of forecast error measures, » *World Applied Sciences Journal*, vol. 24, no. 24, pp. 171–176, 2013.
- [93] J. Schmidhuber, « Deep learning, » *Scholarpedia*, vol. 10, no. 11, p. 32832, 2015.
- [94] P. K. Shivaswamy, W. Chu, et M. Jansche, « A support vector approach to censored targets, » dans *Seventh IEEE International Conference on Data Mining (ICDM 2007)*. IEEE, 2007, pp. 655–660.
- [95] G. A. Satten et S. Datta, « The Kaplan-Meier estimator as an inverse probability of censoring weighted average, » *The American Statistician*, vol. 55, no. 3, pp. 207–210, 2001.
- [96] H. Steck, B. Krishnapuram, C. Dehing-Oberije, P. Lambin, et V. C. Raykar, « On ranking in survival analysis : Bounds on the concordance index, » dans *Advances in neural information processing systems*, 2008, pp. 1209–1216.
- [97] S. R. Safavian et D. Landgrebe, « A survey of decision tree classifier methodology, » *IEEE transactions on systems, man, and cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.
- [98] R. Singh et K. Mukhopadhyay, « Survival analysis in clinical trials : Basics and must know areas, » *Perspectives in clinical research*, vol. 2, no. 4, p. 145, 2011.

## BIBLIOGRAPHIE

- [99] S. Smyl, « A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting, » *International Journal of Forecasting*, vol. 36, no. 1, pp. 75–85, 2020.
- [100] A. Sidahmed et K. Rabhi, « La prédiction du diabete en utilisant les algorithmes de machine learning, » Thèse de doctorat, Université Akli Mohand Oulhadje-Bouira, 2020.
- [101] A. J. Smola et B. Schölkopf, « A tutorial on support vector regression, » *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.
- [102] R. H. Shumway et D. S. Stoffer, « Time series analysis and its applications : with R examples, » 2017.
- [103] A. J. Smola, B. Schölkopf, et K.-R. Müller, « The connection between regularization operators and support vector kernels, » *Neural networks*, vol. 11, no. 4, pp. 637–649, 1998.
- [104] A. A. Tsiatis *et al.*, « A large sample study of Cox’s regression model, » *The Annals of Statistics*, vol. 9, no. 1, pp. 93–108, 1981.
- [105] J. W. Taylor, « Short-term electricity demand forecasting using double seasonal exponential smoothing, » *Journal of the Operational Research Society*, vol. 54, no. 8, pp. 799–805, 2003.
- [106] R. Tibshirani, « Regression shrinkage and selection via the lasso, » *Journal of the Royal Statistical Society : Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [107] M. Tableman et J. S. Kim, « Survival analysis using S : analysis of time-to-event data, » 2003.
- [108] S. J. Taylor et B. Letham, « Forecasting at scale, » *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018.
- [109] O. Uspenskaya-Cadoz, C. Alamuri, L. Wang, M. Yang, S. Khinda, Y. Nigmatullina, T. Cao, N. Kayal, M. O’Keefe, et C. Rubel, « Machine Learning Algorithm Helps Identify Non-Diagnosed Prodromal Alzheimer’s Disease Patients in the General Population, » *The journal of prevention of Alzheimer’s disease*, vol. 6, no. 3, pp. 185–191, 2019.

## BIBLIOGRAPHIE

- [110] B. Vinzamuri et C. K. Reddy, « Cox Regression with Correlation Based Regularization for Electronic Health Records, » pp. 757–766, 2013.
- [111] P. J. Verweij et H. C. Van Houwelingen, « Cross-validation in survival analysis, » *Statistics in medicine*, vol. 12, no. 24, pp. 2305–2314, 1993.
- [112] R. H. Wade, « Financial regime change ? » *New Left Review*, vol. 53, pp. 5–21, 2008.
- [113] Z. Wu et P. D. Christofides, « Economic machine-learning-based predictive control of nonlinear systems, » *Mathematics*, vol. 7, no. 6, p. 494, 2019.
- [114] P. Wang, Y. Li, et C. K. Reddy, « Machine learning for survival analysis : A survey, » *ACM Computing Surveys (CSUR)*, vol. 51, no. 6, pp. 1–36, 2019.
- [115] B. Zupan, J. Demšar, M. W. Kattan, J. Beck, et I. Bratko, « Machine learning for survival analysis : a case study on recurrence of prostate cancer, » *Artificial Intelligence in Medicine*, vol. 20, no. 1, pp. 59 – 75, 2000.
- [116] G. P. Zhang, « Time series forecasting using a hybrid ARIMA and neural network model, » *Neurocomputing*, vol. 50, pp. 159–175, 2003.
- [117] J. Zhang, S. Wang, L. Chen, G. Guo, R. Chen, et A. Vanasse, « Time-dependent survival neural network for remaining useful life prediction, » pp. 441–452, 2019.